1970

Effect of random experimental error in vapor-liquid equilibrium data on thermodynamic consistency

Dean LeRoy Ulrichson
Iowa State University

EFFECT OF RANDOM EXPERIMENTAL ERROR IN VAPOR-LIQUID EQUILIBRIUM DATA ON THERMODYNAMIC CONSISTENCY

by

Dean LeRoy Ulrichson

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Chemical Engineering

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1970

TABLE OF CONTENTS

I. INTRODUCTION

The determination of the vapor phase composition that exists in equilibrium with a liquid phase is of fundamental importance in chemical engineering. The continuing improvements in design techniques, e.g., for distillation and absorption equipment, place an increasing emphasis on the accuracy with which the vapor phase composition can be calculated. The theory of vapor-liquid equilibria is not yet sufficiently advanced to provide an accurate method of predicting equilibrium vapor and liquid compositions. Therefore, equilibrium data must be obtained experimentally. Such experimental work with liquid metal systems at high temperatures in the Ames Laboratory has emphasized the need for a means of analyzing the accuracy of these equilibrium data.

Some of the sources of error in vapor-liquid equilibrium data can not be detected experimentally and classical thermodynamics is used to assist in detecting such errors. This is accomplished by applying the Gibbs-Duhem equation to the data and such an application is commonly called a consistency test. The history of the development of consistency tests is an interesting subject because the Gibbs-Duhem equation has been applied in so many different ways and for so many different purposes. However, the limitations on the Gibbs-Duhem equation, in terms of what it can and can not do, appear not to be well documented. Neither is there any apparent agreement among experimenters on how best to use the Gibbs-Duhem equation to detect experimental errors.

The most common application of the Gibbs-Duhem equation is to data where all of the unknown variables (less those calculated by material

balance considerations) are experimentally measured. Such data are commonly obtained in a "circulation still". A "circulation still" is operated by boiling a liquid in contact with its equilibrium vapor while continuously condensing vapor and circulating it to the still. The temperature, pressure, and vapor and liquid compositions are determined when the conditions in the still are no longer changing with time. The potential sources of error associated with this type of experimental apparatus and procedure have been rather well documented. In general, they apply also to other procedures with only minor modifications. These sources of error can be considered as two basically different types: 1. random measurement errors and ; 2. systematic errors.

Both of these types of error are always present to some extent in experimental data. The random measurement errors, e.g., those associated with uncertainties in the temperature, pressure and composition measurements, can be estimated, and to some extent controlled, before the experiments are performed. The systematic errors, however, are a result of unknown errors associated with the experimental procedure and can not be estimated before the data are taken. Systematic errors arise, for example, from a lack of equilibrium between the vapor and liquid or from entrainment of liquid in the vapor space by too rapid boiling. Such errors are effectively detected, at the present time, only by means of a consistency test.

One other source of error arises in applications of the Gibbs-Duhem equation. This equation contains, in addition to the vapor-

liquid equilibrium variables, terms which depend on the heat of mixing or the volume change of mixing. Since these terms must be measured by separate experimental procedures they are often neglected in consistency tests. It is often difficult to estimate the error incurred by neglecting the heat or volume change of mixing and thus a positive conclusion about the accuracy of the data may not be obtained.

The fact that experimentally determined vapor-liquid equilibrium data should not precisely satisfy the Gibbs-Duhem equation has generally been recognized. However, no universally accepted criterion is available for declaring a set of data "accurately measured". The primary purpose of this work is to develop a procedure which will indicate, with some level of probablility, whether a set of vapor-liquid equilibrium data exhibit no more error than that expected from random measurement errors.

The first step in developing such a procedure is to examine the existing consistency tests for ease and accuracy of application. The most promising test, for the present purposes, appears to be the local area test. The approximate law of propagation of error equation is therefore applied to the local area test equation. This results in a formula for the error bounds (or confidence region) which describes the effect of random measurement error on the consistency test. Statistical techniques are then developed which indicate the probability of a given random distribution in the test results arising only from random measurement errors. A set of data with known uncertainties in the measured variables may then be accepted as free from systematic errors if the probability is high, e.g., the chances are greater than 1 in 10, that only random measurement error exists. Conversely, further analysis

may be required if the probability is low, e.g., less than 1 in 20, that only random measurement error exists. Note that the terms consistent or inconsistent have no meaning in this context. A system which is difficult to handle experimentally, e.g., liquid metals at high temperatures or highly corrosive materials, may have large uncertainties in the measured variables and still be free from systematic error. Small systematic errors may be difficult to detect in the presence of large measurements errors but using the same fiducial limits as with very accurate data will yield equal levels of certainty on the conclusions drawn.

Data for which the consistency test result is non random, or random with a low probability of containing only measurement error, may be treated in several ways. The data could, of course, be rejected as being of unacceptably low accuracy. More probably, the effect of neglected terms in the test equation would be investigated or new estimates of the uncertainty in the measured variables would be sought. Procedures for following both these routes are developed, again with an indication of the confidence that can be placed on the end result.

A set of vapor-liquid equilibrium data tested by this procedure and found to be free of systematic error then has associated with it a somewhat authenticated estimate of the uncertainties in the measured variables. These uncertainties can be used with the propagation of error equations to determine the effect of the measurement errors on the end use for which the data were obtained. As an example, equations are developed which indicate the uncertainty in the activity coefficients. The use of these uncertainties to assist in selecting a

suitably accurate representation equation for the activity coefficients is illustrated. Applications to published data are included.

II. THEORY AND PREVIOUS WORK

The most convenient condition to impose for the attainment of equilibrium between two phases is that the fugacity of component i be the same in each phase. This leads to the expressions

$$x_i \gamma_i f_{il}^0 = f_{il} = f_{iv} = y_i \phi_i P$$

from which the activity coefficient can be defined as

$$\gamma_i = y_i \phi_i P / x_i p_i . \tag{1}$$

The nomenclature here is

x_i = mole fraction of component i in the liquid phase

γ_i = activity coefficient of component i in the liquid phase

p_i = vapor pressure of pure component i (units consistent with P, f_{il} and f_{iv})

f_{il} = fugacity of component i in the liquid phase

f_{iv} = fugacity of component i in the vapor phase

y_i = mole fraction of component i in the vapor phase

ϕ_i = fugacity coefficient of component i in the vapor phase

θ_i = pressure correction term for both the vapor and the liquid phases

P = total pressure (units consistent with p_i)

Throughout this report the activity coefficients are based on the standard state of the pure component at the solution temperature and pressure. The use of this reference state gives

$$\gamma_i = 1.0 \text{ at } x_i = 1.0. \tag{2}$$

The material balances on the vapor and liquid phases are

$$\Sigma y_i = 1.0 \tag{3a}$$

and $$\Sigma x_i = 1.0 \tag{3b}$$

where the sums are taken over all of the N components in each phase. These three equations and the definition of Gibbs free energy form the basis of the vapor-liquid equilibrium relationships. The fundamental equations used in testing for thermodynamic consistency are stated here without derivation. The derivations are given by Ibl and Dodge (21) or by Van Ness (54, 55). The notation used by Van Ness (54) is generally followed except that the function Q replaces $\Delta G^E/RT$, i.e.,

$$Q = \Delta G^E/RT = \sum_{i=1}^{N} x_i \ln\gamma_i. \tag{4}$$

Here

Q = dimensionless excess free energy,

ΔG^E = excess free energy,

R = gas constant, and

T = absolute temperature.

The function Q is also related to $\ln\gamma_i$ through the definition of partial molar properties as

$$\ln\gamma_i = Q + \partial Q/\partial x_i - \sum_{k=1}^{N} x_k \partial Q/\partial x_i. \tag{5}$$

The partial derivatives in equation 5 are taken with all mole fractions constant except the one with respect to which the derivative is taken.

The Gibbs-Duhem equation can be written in several forms. The unrestricted form most commonly used is

$$\sum_{i=1}^{N} x_i d\ln\gamma_i + (\Delta H/RT^2)dT - (\Delta V/RT)dP = 0. \tag{6}$$

Here ΔH is the heat of mixing per mole of solution and ΔV is the volume change of mixing. Another form of this equation written for the liquid

phase is

$$\Sigma x_i d\ln f_i + ((H - H^*)/RT^2)dT - (V/RT)dP = 0. \quad (7)$$

This form was used as the starting point by Ljunglin and Van Ness (29) to derive the general coexistence equation for binary systems, viz.,

$$AdP + BdT = (y_1 - x_1)d\ln(\gamma_{1v}/\gamma_{2v}) + ((y_1 - x_1)/y_1y_2)dy_1. \quad (8)$$

Here

$$A = (\Delta V_v + x_1V_{1v} + x_2V_{2v} - V_1)/RT$$

and $$B = -(\Delta H_v + x_1H_{1v} + x_2H_{2v} - H_1)/RT^2$$

where the subscript v refers to the vapor phase. Equation 8 is called the coexistence equation because it relates the vapor and liquid phases that coexist at equilibrium. Klaus (25) has developed a form of this equation for multicomponent systems. A coupled set of N - 1 non-linear first-order partial differential equations result.

These eight equations are the basis for all of the consistency tests discussed in this thesis. The consistency test procedures that have been proposed in the literature differ primarily in the form of equation 6 or 8 that has been deemed convenient. The test procedure developed here is based on an integration of equation 6.

The Gibbs-Duhem equation may be used in two basically different ways to perform a consistency test. The distinction between these two approaches depends primarily on the number of experimentally measured variables used in the test. Historically, the most common method of obtaining vapor-liquid equilibrium data for an N component system has been to measure N-1 vapor phase compositions, N-1 liquid

phase compositions and the temperature and pressure. The system is completely specified by using the material balance equations to determine the remaining two compositions. The Gibbs-Duhem equation is then an auxiliary relation which is applied to the activity coefficients to test the internal accuracy of the data. This method of testing consistency is designated the direct comparison method.

Another method of testing consistency which has been proposed more recently is applied by calculating equilibrium values for one of the variables. The distinguishing feature of this method is that the Gibbs-Duhem equation is used in place of one of the measured variables to complete specification of the system. With a binary system, for example, one might use only the measured temperature, pressure and liquid composition variables. The Gibbs-Duhem equation and the material balance equations are then used as additional relationships to calculate the vapor composition.

The calculated and measured vapor compositions are then compared as an indication of the accuracy of the data. This technique was developed primarily to provide a procedure for calculating the vapor composition so that the often difficult measurements could be avoided. In that case, the thermodynamic agreement is forced and there is no way to test the data for accuracy. The interest here is only in the consistency test so that all variables are assumed to be experimentally measured.

The test methods described above are based on equations 6 and 8 in which the primary variables are the activity coefficients, the heat of mixing and the volume change of mixing. Another test equation,

proposed by Tao (48,50,51) and later by Lee and Edmister (27) is written in terms of fugacities and the phase enthalpy differences. The uncertainties associated with calculating the derivative terms required in this equation may be quite large. The equation therefore appears to be of limited value as a means of testing the accuracy of vapor-liquid equilibrium data. The method is not discussed in this work.

A. Consistency Tests by Direct Comparison

The direct comparison methods of testing consistency are based on the Gibbs-Duhem equation written in the form of equation 6. The equation may be used, however, in three different ways. (1) A set of data may be tested by dividing equation 6 by dx_1 and estimating the derivatives either graphically from smooth curves or numerically from differences. This procedure is normally called a slope test. (2) Equations, usually polynomials in x_1 with unspecified constants, which satisfy equation 6 at constant temperature and pressure, may be used to fit the activity coefficient data. The results of fitting these representation equations to the activity coefficients for each component are then compared as a test of consistency. Several methods of comparing the fitted equations have been proposed and they are collectively called representation equation tests. (3) Equation 6 may be integrated numerically over a given composition range and the deviation of the result from zero is taken as a measure of consistency. The integral tests compare the areas under the γ_i versus x_i curves and are appropriately named area tests.

1. Slope tests

The earlier slope tests (2) used the Duhem-Margules equation for binary systems in the form

$$x_1 d\ln p_1/dx_1 = x_2 d\ln p_2/dx_2. \quad (9)$$

This equation assumes that either ΔH or ΔV are zero and that the vapor is an ideal gas. Later application of this method used the form (42)

$$x_1 d\ln\gamma_1/dx_1 + x_2 d\ln\gamma_2/dx_1 = 0. \quad (10)$$

This equation does not necessarily depend on the assumption of an ideal gas. The procedure used is to plot $\ln\gamma_1$ and $\ln\gamma_2$ against x_1 and measure the slopes graphically. Because of the low accuracy associated with measuring slopes these equations have not been used extensively. However, they do indicate several useful conditions which can be used to identify grossly inaccurate data. Several sources list these conditions (16,30, 42, 54).

Slope tests have also been developed that require measuring the slope of the Q versus liquid composition curve. This method is discussed by Van Ness and Mrazek (58), Van Ness (54, 57) and Prausnitz (37) and has been extended to ternary systems by Lu and Chang (31). The general procedure is to plot Q versus x_1 and draw a tangent line to the smooth curve at a point b. The intersection of the tangent line with the ordinate at $x_1 = 1$ gives a value of $\ln\gamma_1$ at $x_1 = b$. The accuracy of this method depends on the curvature of the Q versus x_1 curve and the length of the extrapolation required for large and small values of x_1. An alleged improvement, called the composition-resolution method (58), is to use a plot of Q/x_1x_2 versus x_1. The curvature with Q/x_1x_2 is

usually smaller than with Q and the values of $\ln\gamma_i$ are obtained from the opposite intercepts so that the length of extrapolation is shorter. In addition, the extension of the Q/x_1x_2 curve to $x_1 = 0$ and $x_1 = 1$ gives the activity coefficients at infinite dilution directly. That is, at $x_1 = 0$, $Q/x_1x_2 = \ln\gamma_1$ and at $x_1 = 1$, $Q/x_1x_2 = \ln\gamma_2$ as can be verified from equation 4.

The equations for the composition-resolution test are derived with reference to Figure 1. From the definition of a tangent at $x_1 = b$

$$(c - I_1)/b = \partial(Q/x_1x_2)/\partial x_1.$$

Solving for I_1 and substituting x_1 for b and Q/x_1x_2 for c gives

$$I_1 = Q/x_1x_2 - x_1\partial(Q/x_1x_2)/\partial x_1. \tag{11}$$

Evaluating the partial derivative by substitution of equation 4 gives

$$\partial(Q/x_1x_2)/\partial x_1 = (-(x_2/x_1)Q + (Q + x_2\partial Q/\partial x_1))/x_1x_2^2 \tag{12}$$

Applying equation 5 to a binary system yields

$$\ln\gamma_1 = Q + x_2\partial Q/\partial x_1 \tag{13}$$

and combination of 11, 12 and 13 gives, after rearranging,

$$\ln\gamma_1 = x_2^2\ (2Q/x_1x_2 - I_1). \tag{14a}$$

Similarly

$$\ln\gamma_2 = x_1^2\ (2Q/x_1x_2 - I_2). \tag{14b}$$

This test has received considerable attention in recent years. Van Ness (57) has stated, however, that this is a very exacting test and that random errors must be small to establish a smooth curve for Q/x_1x_2 versus x_1. He also points out that it is only a positive test. That is, if the experimental and derived values do not agree, the data can only be rejected as inconsistent. This point will be discussed later in some detail.

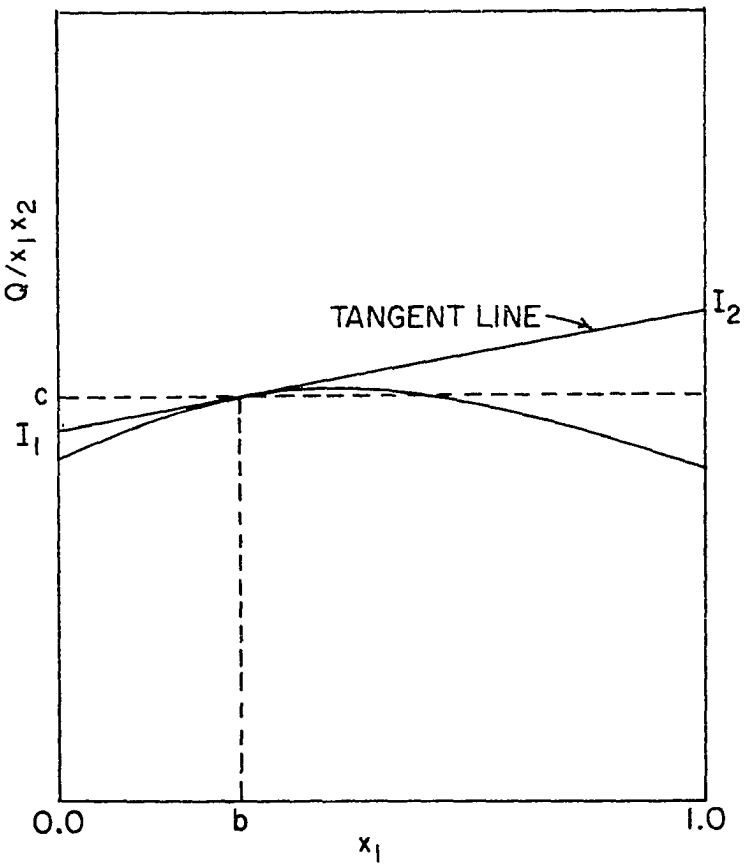

Figure 1. Illustration of composition-resolution test

Another slope test which requires a rather complex regression analysis is described by Techo (52). His method is developed for binary isothermal data where the volume change of mixing is negligible. The excess chemical potentials are written in terms of virial coefficients as

$$RT\ln\gamma_1 = RT\ln(Py_1/p_1x_1) + (B_{11} - V_1)(P - p_1) + \beta_{12}Py_2^2 \qquad (15)$$

with a similar expression for component 2. In equation 15 B and β are virial coefficients and V_1 is the liquid volume of pure component one. Partial differentiation of these expressions and substitutuion into equation 10 gives, after rearranging,

$$RT/P - (B_{11} - v_1)x_1 + (B_{22} - v_2)x_2 + \beta_{12}(y_2^2x_1 - y_1^2x_2))(dP/dy_1) +$$
$$(RT((x_1 - y_1)/y_1y_2) + 2\beta_{12}P(y_1 - x_1)) = 0. \qquad (16)$$

All the terms in equation 16 are experimentally determined quantities except the derivative dP/dy_1, and the virial coefficients. Several methods of estimating virial coefficients are available. Techo chose to evaluate the derivative by determining $P = P(x)$ and $y = y(x)$ in terms of the orthogonal polynomials described by Forsythe (10). The order of the polynomials $P(x)$ and $y(x)$ is automatically determined by requiring that the deviation from the null value in equation 16 be a minimum. Polynomials of order 5 to 15 are typical in applying this procedure. The consistency test is then simply a comparison of the deviation of equation 16 from the null value. Techo pointed out that, since both P and y are fixed at $x_1 = 0$ and 1, the consistency of the data can be checked as near the endpoints as desired. The method therefore has a significant advantage over slope tests which use the activity coefficients directly. However, he did not indicate whether

any extraneous inflection points occurred in using such a high polynomial order. Klaus (25), for example, found inflection points occurred for most systems with 5th or 6th order polynomials. The existence of extraneous inflection points in the polynomials for P and y could give a significant error in the derivative term in equation 16 and, therefore, false indications of consistency or inconsistency. Techo also proposed a maximum limit for accurately measured data in terms of the sums of the squares of the deviations obtained from equation 16.

The limiting slopes of P, T, y and γ curves with respect to liquid composition are often of interest in correlating, predicting or testing vapor-liquid equilibrium data. Several papers have dealt with this subject, e.g., (12, 20, 30).

2. Representation equation tests

Many representation equations, which are solutions of equation 10, have been presented in the literature. Some of the more frequently used equations are the Margules, Van Laar and Scatchard-Hamer equations, usually used in the form presented by Carlson and Colburn (7), and the Redlich-Kister equations (41). An equation more recently proposed by Wilson (60) appears to be very popular because it has some built-in temperature dependence which the others do not.

Representation equations have been used to smooth activity coefficient data and to test for thermodynamic consistency. However, in order to smooth only the random error from activity coefficient data, an equation capable of representing the liquid phase behavior exactly must be available. There have been some advances in relating the constants in

representation equations to the chemical properties of the liquid phase but there is no sure way to choose the right equation for a given set of data. There is always doubt, therefore, whether any lack of fit is due to the inaccuracy of the data or inadequacy of the representation equation. Despite this fact, representation equations have often been used to test thermodynamic consistency.

From an examination of the Margules and Van Laar equations Carlson and Colburn (7) pointed out several useful qualitative checks on activity coefficient data. Lu (30) listed eleven quick tests which are based on equation 10 and representation equations. Tierney (53) used a statistical analysis to analyse the random error from a fit of the Van Laar or Margules equations.

Redlich and Kister (42) proposed using a representation equation in the form

$$Q = x(1 - x)(a + b(2x - 1) + c(2x - 1)^2 + ---). \qquad (17)$$

They then suggested plotting $\ln\gamma_1/\gamma_2$ against x_1 for binary data and determining constants in equation 17 from

$$dQ/dx_1 = \ln\gamma_1/\gamma_2 = a(1 - 2x) + b(6x(1 - x)-1) + c(1 - 2x)(1 - 8x(1 - x) + ---. \qquad (18)$$

Using this equation and selected values of x_1 permits solving directly for a, b and c. Seven useful values of x_1 were suggested for solving for the constants. They also suggested fitting Q/x_1x_2 by least squares if numerous data points were available. Jost and Röck (22) extended this concept by using Legendre orthogonal polynomials and Vettins discontinuous orthogonal polynomials. Gilmont, et al., (13) have

also used this type of approach to represent the relative volatility and thus the activity coefficients. They also obtain a criterion for thermodynamic consistency from the relationship between the two components.

Probably the most elegant mathematical treatment of representation equations is that due to Klaus (25). Klaus and Van Ness (23, 24) presented two sets of equations, however, because the orthogonal polynomials which they considered first would not satisfactorily represent the shapes of all the types of curves encountered. This is, in itself, an indication of the problems associated with attempting to use a representation equation to determine thermodynamic consistency. The procedure used by Klaus in developing the orthogonal polynomial representation is briefly discussed.

The orthogonal polynomial set described by Forsythe (10) was used with some modifications. Although Klaus developed a procedure whereby any two of the three general thermodynamic quantities M, M_1 or M_2 ($M = x_1M_1 + x_2M_2$) could be represented, these quantities will be taken to be Q, $\ln\gamma_1$ and $\ln\gamma_2$, respectively, for the present discussion. The orthogonal polynomials were developed in such a manner that if $\ln\gamma_1$ and $\ln\gamma_2$ were fit independently, a consistency test could be applied and the coefficients of the series could be modified to make the data consistent. Two methods of modifying the coefficients were presented. One requires a simple average of the coefficients of each series while the other was developed by applying an analytical form of the composition-resolution test. Klaus shows that the composition-resolution method makes the data consistent while minimizing the shift in the value of Q. Conversely the averaging technique minimizes the shift

in the partial property curves--$\ln\gamma_i$ in this case. In all the examples presented, the averaging technique appeared to give the best result.

3. Area tests

The consistency tests most discussed in the literature are based on an integration of equations 6 and 10. Redlich and Kister (41) and Herington (18) independently presented what is called the area test in 1947 and 1948 in the form

$$\int_0^1 \ln\gamma_1/\gamma_2 dx_1 = 0. \tag{19}$$

Equation 19 is obtained by integrating equation 10 by parts and noting that $Q(1) = Q(0) = 0$ and $dx_1 = -dx_2$. This test is the one most commonly used by authors to indicate the consistency of new data which they are presenting.

The name given to this test results from the fact that the $\ln(\gamma_1/\gamma_2)$ versus x_1 curve passes through $\ln(\gamma_1/\gamma_2) = 0$ and the area above the line $\ln(\gamma_1/\gamma_2) = 0$ must equal the area below the line for the data to be consistent. This area difference is often reported as a measure of consistency. Redlich, Kister and Turnquist (43) later showed that the area test could not distinguish between two sets of data taken by different experimenters even though the data differed significantly. The reason was that the areas above and below the line compensated so that the value of the integral was the same in each case. Techo (52), Van Ness (57) and Bourne (5) have commented on the deficiencies in this test. Probably the most serious criticisms are that data are required over the whole range of composition, compensating errors are not detected, and for binary data the temperature and

pressure can not both be held constant while varying composition. Both Redlich-Kister and Herington noted the latter problem and Herington (17) presented criteria for deciding whether the heat of mixing term in equation 6 could be ignored for isobaric data. Thus he developed criteria which would indicate whether a nonzero value in equation 19 could be considered consistent or not without integrating the heat of mixing term of equation 6.

The use of the area test for binary systems was further considered by Broughton and Brearley (6), Black (3), and Black, Derr and Papadopoulos (4). Broughton and Brearley assumed that the product $RT\ln\gamma$ was independent of temperature and that the heat of mixing term could be neglected by integrating this product for isobaric data. Van Ness (57), among others, has pointed out that this is not a safe assumption. Black and Black, Derr and Papadopoulos, in a discussion of representation equations and consistency tests, indicated that the area test was insensitive to temperature and total pressure measurement errors. Black therefore proposed (based on an analysis of his representation equations) that a smooth plot of $\sqrt{\ln\gamma_1}$ versus $\sqrt{\ln\gamma_2}$ would indicate accurate pressure or temperature measurements and that simultaneous satisfaction of the area test would indicate consistent data. The area test is insensitive to errors in temperature and total pressure because $\ln\gamma_1/\gamma_2 = \ln((y_1P/x_1p_1)(x_2p_2/y_2P))$ so that total pressure cancels and temperature errors appear as a ratio through vapor pressures.

Other extensions of the integral tests made use of short intervals of integration. These extensions considered tests for both binary and

ternary systems. For the ternary and multicomponent systems some particular path of integration must be chosen to integrate equation 20, which is obtained from equation 6 by integration by parts.

$$Q(b) - Q(a) - \sum_{i=1}^{N} \int_a^b \ln\gamma_i dx_i + \int_{T_a}^{T_b} (\Delta H/RT^2)dT - \int_{P_a}^{P_b} (\Delta V/RT)dP = 0. \quad (20)$$

Herington (15, 17) considered integration along a path where the ratio of two mole fractions was a constant for conditions where the temperature and pressure terms were negligible. Krishnamurty and Rao (26) proposed integration with one mole fraction constant for negligible temperature and pressure terms. Li and Lu (28) considered short linear paths chosen such that T and P were nearly constant. They also suggested integrating over a closed loop. Prausnitz and Snider (38) presented a general method for multicomponent systems by holding x_i/x_1 constant for i = 2,3,---N - 1. Mc Dermott and Ellis (33) modified the method of Li and Lu by using the trapezoid rule to compare points by pairs. All of these methods are similar except that different paths of integration are chosen. However, the result of applying equation 20 over some interval a, b is not expected to be zero even if only random error is present in the data. A criterion for relating the deviation of equation 20 from zero to the accuracy of the data was not advanced. Even a rigorous application of equation 20 is of doubtful value unless some meaning can be attached to the result.

The meaning of the area tests can be shown graphically by integrating equation 10 over a short interval b - a. For a binary system the integral is

$$\int_a^b x_1 d\ln\gamma_1 + \int_a^b x_2 d\ln\gamma_2 = 0. \quad (21)$$

This equation implies that the strips indicated in Figure 2 must be equal for each interval b - a. Obviously the limits on the integrals refer to $\ln\gamma_i$ evaluated at $x_1 = a$ and $x_1 = b$. The limits of a = 0 and b = 1 result in another form of equation 19. Applying these limits in Figure 2 indicates that the total areas under each curve must also be equal. This fact has led Van Ness (57) to suggest that the two areas be considered average ordinates and the inconsistency be reported as an error in the average ordinate. Note that for a fixed $\ln\gamma_1$ versus x_1 curve there are many $\ln\gamma_2$ versus x_1 curves which will yield the same total area for a = 0 and b = 1. However, requiring small area segments under each curve, e.g., b - a = 0.05, to be equal uniquely fixes the relationship between the curves.

Activity coefficient data are often obtained for systems which contain a nonvolatile component. The area test can then be rearranged and used to calculate one activity coefficient from a knowledge of the other, viz.,

$$\ln\gamma_2 = -\int_{\ln\gamma_{1a}}^{\ln\gamma_{1b}} (x_1/x_2)\, d\ln\gamma_1. \qquad (22)$$

This equation could also be used as a consistency test if the $\ln\gamma_2$ data were known. The problem with this approach is that an error in the $\ln\gamma_1$ data at low x_1 is carried through the whole composition range and does not give a true pointwise comparison at each of the $\ln\gamma_2$ data points.

The problem of propagation of error from low x_1 in equation 22 led Tao (46, 49) to propose using a combination of the slope and integral tests. His method also included an analysis of the effect of random experimental error.

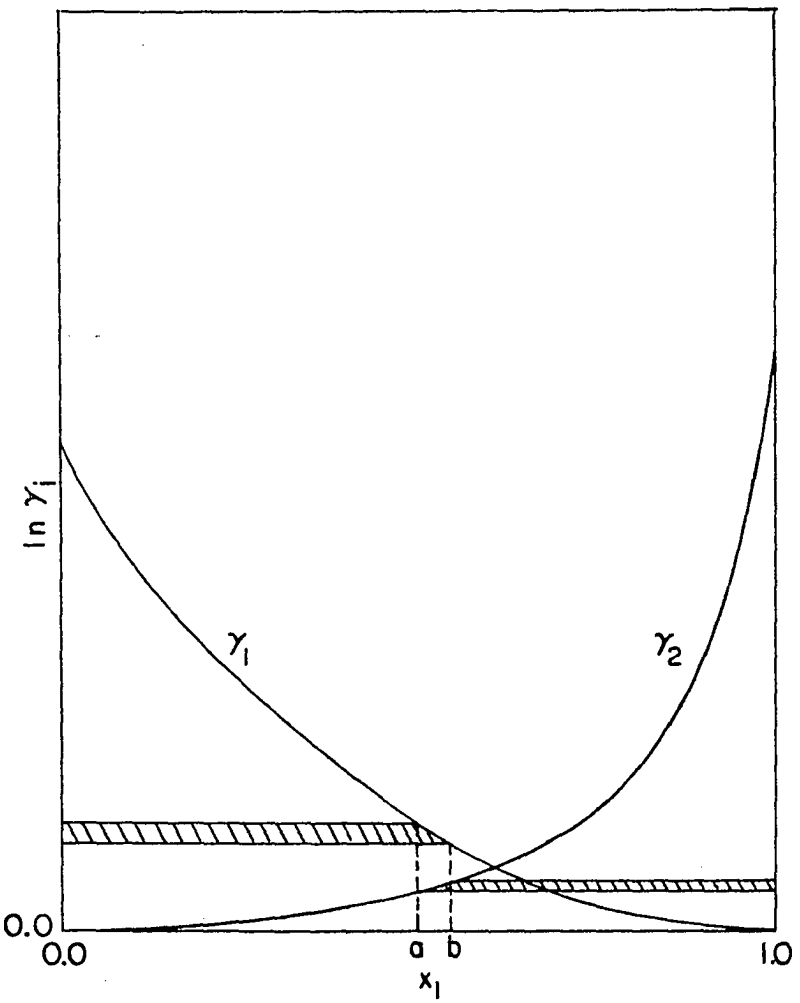

Figure 2. Illustration of local area test

Tao defined the quantities Q and z as

$$Q = \Sigma x_i \ln\gamma_i \tag{4}$$

$$z = A(x_1) + B(x_1) + \Sigma K_i \ln\gamma_i, \tag{23}$$

where

$$A(x_1) = -(\Delta H/RT^2)\, dT/dx_1$$

$$B(x_1) = (\Delta V/RT) dP/dx_1$$

and

$$K_i = dx_i/dx_1 = (x_{ia} - x_{ib})/(x_{1a} - x_{1b}).$$

Then, since z and Q can both be obtained directly from the experimental data and are also related as

$$z = dQ/dx_1 \tag{24}$$

and

$$Q(x_{1b}) - Q(x_{1a}) = \int_{x_{1a}}^{x_{1b}} z\, dx_1, \tag{25}$$

a cross-check on the data is possible. Both z and Q are calculated from experimental data and plotted against x_1. A second set of z values is then obtained from equation 24 and a second set of Q values is obtained from equation 25. The difference between these two sets of z and Q values constitutes the consistency test. Since the two sets will never agree precisely Tao added a noise bound for each curve by using total derivatives of equations 1, 4 and 23 in the forms

$$E(\gamma_i)/\gamma_i = E(P)/P + (E(T)/p_i) dp_i/dT + (1/y_i + 1/x_i)E(x), \tag{26}$$

$$E(Q) = \Sigma(x_i E(\gamma_i)/\gamma_i - E(x)\ln\gamma_i), \tag{27}$$

and

$$E(z) = \Sigma K_i E(\gamma_i)/\gamma_i, \tag{28}$$

where, for example, $\pm E(w)$ represents the maximum error in w. If the

differences between the two sets of values of Q and z fall within the error bands generated from 27 and 28 the data are considered consistent. If they do not, and A, B or vapor phase nonideality have not been evaluated, Tao suggests procedures for estimating whether these terms will make the data consistent. Thus decisions of consistency, inconsistency or conditional consistency are reached.

Stevenson and Sater (45) later suggested applying equation 20 directly. They pointed out that since Tao had chosen $x_{1a} = 0$ in his examples, an extrapolation error would often be included in 25 as discussed for equation 22. Although the slope test would theoretically indicate extrapolation errors, slopes cannot be measured as precisely as might be desired. Stevenson and Sater set equation 20 equal to a function, $f(x_1)$, instead of zero and then plotted $f(x_1)$ against x_1 for binary systems. The deviation of $f(x_1)$ from zero then indicates the inaccuracy (or inconsistency) of the data. They also demonstrated that if $f(x_1)$ was cumulated from $x_1 = 0$, much of the sensitivity to local inconsistency was lost. Although their method is applicable to multicomponent paths by appropriate choice of integration paths, they illustrated the test only for binary systems.

Chang and Lu(8) combined the techniques proposed by Tao (49) and Stevenson and Sater (45) to develop maximum error bounds for a local area test. Their method is based on equation 10 with the activity coefficients corrected to a reference temperature and pressure. Equation 10 is then rigorous and, after integrating by means of the trapezoid rule, is written as

$$\sum_{i=1}^{N} (x_{ia} + x_{ib})(\ln\gamma_{ib} - \ln\gamma_{ia}) = 0. \tag{29}$$

The subscripts a and b are used to indicate adjacent data points. The equation is applied between each pair of adjacent data points to be tested. Substituting the definition of the activity coefficients in equation 29, the proposed test equation becomes

$$d = \sum_{i=1}^{N} (x_{ia} + x_{ib})(\ln(K_{ib}/K_{ia}) - \ln((\phi_{ia}P_a)/(\phi_{ib}P_b)) - \int_{P_a}^{P_b}(V_{il}/RT)dP - \int_{T_a}^{T_b}((h_{iv} - \bar{H}_{il})/RT^2)dT. \tag{30}$$

The terms in equation 30 are defined as

$K_i = y_i/x_i$,

ϕ_i = fugacity coefficient of component i in the vapor mixture,

$\bar{V}_{il}$ = partial molar liquid volume of component i,

h_{iv} = ideal gas enthalpy of component i in the vapor state, and

$\bar{H}_{il}$ = partial molar liquid enthalpy of component i in the solution.

The integral terms involving volume and enthalpy arise in this form because of the correction to a reference temperature and pressure. When testing isothermal or isobaric data one of these integrals will obviously vanish.

Chang and Lu then developed maximum error bounds for d by assuming the supporting quantities--volume, enthalpy and fugacity coefficients--to be free from error. The total derivative of d, with $dx_i = dy_i$, dP and dT replaced by $\pm E(x)$, $\pm E(P)$ and $\pm E(T)$, respectively, then yields the following expression for the maximum error bound under isothermal conditions.

$$D = \sum_{i=1}^{N} \{(x_{ib} + x_{ia})(1/y_{ib} + 1/y_{ia} + 1/x_{ib} + 1/x_{ia}) +$$

$$2(\ln(K_{ib}/K_{ia}) - \ln(\phi_{ia}P_a/\phi_{ib}P_b) - \int_{P_a}^{P_b}(\bar{V}_{il}/RT)dP)\}E(x) +$$

$$\sum_{i=1}^{N}(x_{ib} + x_{ia})\{1/P_a + 1/P_b + (\partial/\partial P\int_{P_a}^{P_b}(\bar{V}_{il}/RT)dP)\}E(P) +$$

$$\sum_{i=1}^{N}(x_{ib} + x_{ia})(\partial/\partial T\int_{P_a}^{P_b}(\bar{V}_{il}/RT)dP)E(T). \qquad (31)$$

The maximum error bound for isobaric conditions is developed similarly to give

$$D = \sum_{i=1}^{N} \{(x_{ib} + x_{ia})(1/y_{ib} + 1/y_{ia} + 1/x_{ib} + 1/x_{ia}) +$$

$$2(\ln(K_{ib}/K_{ia}) - \ln(\phi_{ia}/\phi_{ib}) - \int_{T_a}^{T_b}((h_{iv} - \bar{H}_{il})/RT^2)dT)\}E(x) +$$

$$\sum_{i=1}^{N} 2(x_{ib} + x_{ia})E(P)/P +$$

$$\sum_{i=1}^{N}(x_{ib} + x_{ia})(\partial/\partial T\int_{T_a}^{T_b}((h_{iv} - \bar{H}_{il})/RT^2)dT)E(T). \qquad (32)$$

In equations 31 and 32 the absolute values are to be taken for all the expressions enclosed in parentheses. The term E(x) represents the maximum measurement errors in both x and y, and E(P) and E(T) represent the maximum measurement errors in P and T, respectively. According to Chang and Lu, an experimental point is considered inconsistent if the value of d calculated for the pair containing that point is greater in absolute value than the experimental error bound D. They illustrated the application of equations 30, 31, and 32 by testing both binary and ternary systems. The systems tested contained components of air and natural gas for which the notation used was particularly convenient. This test would appear to be a particularly good one except that the term "maximum error bounds" is quite vague and the errors in x and y

are not considered separately. Also the maximum measurement error in P, for example, may be much greater than the average measurement error. Translating the maximum measurement error in P to the error bounds, D, may then obscure small systematic errors. The use of maximum errors also prevents application of statistical criterion to reach a conclusion about the accuracy of the data.

Other discussions of the integral test have been given by Deshpande and Lu (9), Lu and Jones (32) and Herington (18). Deshpande and Lu and Lu and Jones used integration procedures similar to those suggested for equation 22 to extrapolate data over the full concentration range from a few points. Herington pointed out some interesting symmetrical properties of the integral of equation 10 and discussed the sensitivity of such a test to various kinds of measurement errors. Although some of the sensitivities have been previously discussed by Black (3) and Bourne (5), they did not consider the symmetrical tests proposed by Herington. He shows that the integral of equation 10 can be written as

$$\int_0^b x_1 d\ln\gamma_1 + \int_b^1 x_2 d\ln\gamma_2 = \int_0^1 x_1 d\ln\gamma_1 = \int_0^1 x_2 d\ln\gamma_2, \qquad (33)$$

where the integral limits mean the value of x_1 at which $\ln\gamma_i$ is evaluated. Referring to the first two terms on the left in equation 33 as I_1 and I_2 respectively, Herington shows that $I_1 + I_2$ = constant. Nine properties of the functions I_1 and I_2 are then given and proposed as tests. The only test which can not as well be applied to equation 10 directly is the plot of $I_1 + I_2$ versus x_1 which should be a straight line. However, the purpose of such a plot is to locate the composition range in which errors may have occurred. The procedure suggested by Stevenson

and Sater (45) in plotting $f(x_1)$ against x_1 accomplishes this purpose much more efficiently.

B. Calculation of Equilibrium Values

The methods of calculating consistent vapor phase compositions and liquid phase activity coefficients are of two general classes:

1. those methods based on equation 8, the coexistence equation, and
2. those based on combinations of equation 1 through 6, or indirect methods.

1. Coexistence equation

The coexistence equation is given by Ljunglin and Van Ness (29) as equation 8. However, in its early applications it was not presented in a rigorous fashion. Redlich and Kister (42, 43) proposed its use as a slope test which could be used in conjunction with the area test. They derived the equation

$$dT/dy_1 = s(x_1 - y_1)/y_1y_2 \tag{34}$$

where

$$1/s = x_1 d\ln P_1/dT + (1 - x_1) d\ln P_2/dT, \tag{35-36}$$

and pointed out that RT^2/s represents the ideal heat of vaporization for one mole of solution. The function P_1 and P_2 are defined to include temperature corrections and non-ideal vapor phase corrections. Redlich and Kister (42) indicated that the most desirable method of checking thermodynamic data was to develop a relationship between x and y. Material balances and consistency were then automatically satisfied and the derived functions γ_1 and γ_2 could be calculated from the consistent data. This kept the analysis close to the experimental data rather

than putting all experimental errors into the derived quantities γ_1 and γ_2. They demonstrated the use of this equation in the form of equation 34 by plotting T versus x and y and calculating slopes through each T, y data point. The data points were then shifted by trial and error to permit drawing a smooth curve which agreed with the slopes at each point.

In a later paper Redlich, Kister and Turnquist (43) discussed the use of

$$d\ln P/dy_1 = (y_1 - x_1)/y_1y_2 \qquad (37)$$

to test data taken at constant pressures. Ljunglin and Van Ness (29) discussed the assumptions required to reduce equation 8 to equations 34 and 37. Several discussions of these equations have appeared in the literature.

Othmer, Ricciardi and Thakar (36) presented a more straightforward and rigorous derivation of equation 34 by assuming only that the Clausius-Clapeyron equation and Dalton's law of partial pressure applied. They suggested graphical integration of the equation as well as demonstrating its use to test consistency. Ho, Boshko and Lu (19) suggested the integration of equation 37 but indicated that the method could not be applied successfully to azeotropic systems because of difficulties in fitting the azeotrope portion of the P-x curve with a polynomial.

Ljunglin and Van Ness (29) presented a rigorous derivation of equation 8 and discussed various methods of approximating the terms A and B. They presented a solution of the equation for both constant temperature and constant pressure data by using the second virial

coefficient to represent vapor phase nonidealities. They did not discuss the method used to obtain equal interval P-x data required for the integration although they did study an azeotropic system. Apparently, the azeotrope composition was determined from the x-y curve after integrating nearly to the azeotrope from x = 0 and from x = 1. Klaus (25), using virial coefficients in the same manner, derived equations which relate the coexistence of vapor and liquid for multicomponent systems. The solution of these equations was not attempted however.

Lu and Chang (31) have applied equation 8 to a ternary system by treating the ternary as a number of binaries where one of the binary components is actually a mixture of two components. The excess total pressure, defined as $P^E = P - x_1p_1 - x_2p_2$ was used in developing a polynomial for pressure in terms of composition. The system did not exhibit an azeotrope and excellent results were achieved. Note that this procedure requires no assumptions about ternary interaction.

Several articles have considered the difficulties encountered in integrating equation 8, particularly with respect to azeotropic systems. Two of the most recent are by White and Lawson (59) and Van Ness (56). They indicate that the correct solution cannot be obtained unless the proper starting point is used for the integration. Selection of the proper starting point appears to be particularly important for azeotropic systems.

2. Indirect methods

The indirect methods, as classified here, generally depend on a

series of successive approximations to a total pressure or excess free energy surface. These methods have been developed primarily as procedures for calculating equilibrium phase compositions from P - x or T - x data. That is, from data where the vapor phase composition is not experimentally measured.

One of the first of these methods is discussed by Barker (1). Polynomial expressions are used to describe both the total pressure and the dimensionless excess free energy, Q. Thus the final result for the excess free energy contains both the errors in the representation of the total pressure and the errors associated with measuring the P - x data. Modifications of this procedure have been proposed by Prengle and Palm (39) and Prengle and Pike (40).

Another procedure for calculating vapor phase compositions from total pressure methods has been proposed by Mixon, Gumowski and Carpenter (34). They also present an interesting discussion of the relative merits of the available procedures for calculating vapor-liquid equilibrium conditions from total pressure measurements. They point out that the coexistence equation becomes quite complex for multicomponent systems and that they have been unable to obtain meaningful solutions for the highly nonideal acetone-chloroform system. They criticize Barker's method because a functional form for the excess free energy must be initially assumed whereas the coexistence equation and their method require no assumptions about the liquid phase behavior.

A unique method proposed by Tao (47) appears to be completely rigorous but it does not generalize to multicomponent systems. One advantage of Tao's method that the above methods do not have is that it

can be used for either P - x or T - x data. The only requirement being that a method of smoothing and interpolating be available for either the P - x or T - x data. The major objection to this method is that the activity coefficient for component one must be known at $x_1 = 0$. This information may be estimated from the equations derived by Gautreaux and Coates (12) but the accuracy is likely to be low.

The final test method to be discussed is one proposed by Friend, Scheller and Weber (11). The method is presented for binary systems although generalization to multicomponent systems may be possible. Their procedure requires that experimental values for x_1/y_1, P and T be known. One of the four variables is eliminated by using the Gibbs-Duhem equation and thermodynamically consistent values for that variable are calculated from the other three. Another variable is then eliminated and the procedure repeated until all four variables have been calculated. The effect of this technique is to assume that all of the error is in the variable being calculated. Therefore, maximum uncertainties are obtained for each variable. Unfortunately, the procedure requires cumulative integration, by means of the trapezoid rule, over the complete range of the variable. This means that errors arising from the integration may become significant. The cumulative integration errors are, therefore, difficult to distinguish from systematic errors. The authors did not suggest any statistical tests for determining the significance of the deviations obtained for each variable.

III. RESULTS

A. Comparison of Consistency Tests

The primary goal of this work was to develop a rigorous procedure for testing vapor-liquid equilibrium data. The procedure should contain statistical techniques which characterize the random measurement errors and also detect systematic errors. The test must, therefore, be based on methods for which the errors involved in application, e.g., numerical integration errors, can be estimated. Since the errors involved in graphical applications of consistency tests are difficult to estimate, graphical tests are not considered in this work. The classification by type is followed in discussing the applicability of the available tests.

1. Consistency tests by direct comparison

a. Slope tests Slope tests are based on equation 6 or 10. Defining a deviation function for the slope test, as has been done for the local area test, these equations may be written as

$$E_k = \sum_{i=1}^{N} x_i (d\ln\gamma_i/dx_k) + L. \qquad (38)$$

Here L is defined as $(\Delta H/RT^2)dT/dx_k$ for isobaric data and as $-(\Delta V/RT)dP/dx_k$ for isothermal data. E_k is the error which arises in the Gibbs-Duhem equation with respect to component k. Thus a nonzero value of E_k is a measure of the uncertainties associated with the data. The slope test has not generally been used in this differential form because of the inherent inaccuracy associated with calculating derivatives numerically. However, even if the derivatives could be satisfactorily estimated, a means of determining acceptable bounds for E_k has not been advanced. One method of

characterizing the effect of E_k on the accuracy of the activity coefficients is developed here by using equation 5.

Equation 5 as written contains one dependent variable and can not be used for direct application to data since the material balance condition 3b must also be met. The dependent variable can be eliminated by writing 5 as

$$\ln\gamma_i = Q - \Sigma x_k \partial Q/\partial x_k, k \neq i. \tag{39}$$

The composition variables are then all independent and equation 4 can be substituted and the indicated operations performed. All mole fractions are held constant in differentiation except x_k and x_i. The result is

$$\begin{aligned}\ln\gamma_i &= Q - \Sigma x_k(\ln\gamma_k - \ln\gamma_i + \Sigma x_i \partial \ln\gamma_i/\partial x_k + L) \\ &= Q - \Sigma x_k \ln\gamma_k + (1 - x_i)\ln\gamma_i - \Sigma x_k(\Sigma x_i \partial \ln\gamma_i/\partial x_k + L) \\ &= Q - \Sigma x_k \ln\gamma_k - x_i \ln\gamma_i + \ln\gamma_i - \Sigma x_k E_k \\ &= \ln\gamma_i - \Sigma x_k E_k \end{aligned} \tag{40}$$

where all summations on k are from 1 to N with component i excluded. For consistent data, i.e., $E_k = 0$, this obviously reduces to an identity.

Equation 40 can provide an indication of the error in γ_i. If E_k is determined from equation 38, γ_i is taken as a data point and γ_i^c indicates a consistent value of γ_i, then

$$\gamma_i^c = \gamma_i \exp(-\Sigma x_k E_k), \; k \neq i. \tag{41}$$

For a binary system equation 41 can be written as

$$\gamma_1^c = \gamma_1 \exp(-x_2 E_2), \tag{42a}$$

$$\gamma_2^c = \gamma_2 \exp(-x_1 E_1) \tag{42b}$$

with

$$E_1 = x_1 d\ln\gamma_1/dx_i + x_2 d\ln\gamma_2/dx_1 = -E_2.$$

Equations 42a and 42b provide a means of calculating a set of activity coefficients, γ_1^c and γ_2^c, which are thermodynamically consistent. It can also be shown that they are partial molal properties of the same dimensionless excess free energy as would be obtained from the experimental activity coefficients. Taking logarithms of equations 42a and 42b, multiplying 42a by x_1, 42b by x_2, and adding the result, gives

$$x_1 \ln\gamma_1^c + x_2 \ln\gamma_2^c = x_1 \ln\gamma_1 + x_2 \ln\gamma_2 - x_1 x_2 E_2 - x_1 x_2 E_1.$$

Since $E_1 = -E_2$ this reduces to an identity in Q. The consistent set of activity coefficients is therefore obtained by changing the value of the partial molal properties, $\ln\gamma_i$, while holding the total property, Q, constant. The value of Q may have great theoretical significance but it is not a measured quantity. Therefore, if the activity coefficients are incorrect, the value of Q is also likely to be incorrect and the incentive for calculating the consistent set of activity coefficients is lost.

Equation 42a and 42b can also be obtained by manipulation of the composition-resolution test equations and explains why the composition-resolution test is only a positive test. That is, when γ_i^c is not equal to γ_i the data can only be rejected as incorrect; the new set of values should not be used. Equation 41 can apparently be considered a form of the multicomponent version of the tangent-intercept or composition-resolution tests. The evaluation of the partial derivatives which arise for multicomponent systems may be somewhat difficult, however.

The slope tests do not appear to provide the type of quantification

of a consistency test that is presently desired. The representation equation tests are therefore examined.

b. Representation equation tests Many different forms of representation equations have been presented in the literature along with discussions of their relative merits. The Margules equations will be used here primarily because they are simple. They illustrate some of the possible uses and some of the defects in this approach.

The use of a prespecified form of representation equation has been criticized because this is equivalent to preselecting a model to describe the liquid phase behavior. Several forms of unspecified orthogonal polynomials have been presented which avoid this criticism. The criterion for deciding what order of polynomial best represents the data becomes critical, however, for orthogonal functions. The problem is particularly acute if the data are to be differentiated. Extraneous inflection points must then be avoided and some systems have been found which can not be even approximately represented by such functions without introducing unwarranted inflection points (24). Neither are such systems well approximated by the Margules equations. The use of representation equations to test thermodynamic consistency thus has one limitation at the outset. Extremely nonideal systems and systems with only a few data points are not amenable to treatment.

Another problem occurs with representation equations. It is impossible to determine whether the lack of fit results from the inadequacy of the equation or from errors in the data. This also makes it important to consider carefully the form of the data to be represented. Equations have been presented which are suggested for use

in fitting Q, Q/x_1x_2, $\ln\gamma_1$, $\ln\gamma_2$, $\ln\gamma_1/x_2$, $\ln\gamma_2/x_1$ and $dQ/dx_1 = \ln\gamma_1/\gamma_2$. Techo (52) found that fitting dQ/dx_1 gave slightly better results than fitting $\ln\gamma_1$ and $\ln\gamma_2$ separately. Redlich and Kister (41, 42) present equations best suited to fitting either Q or dQ/dx_1. Klaus and Van Ness (24) fit two of the three variables Q/x_1x_2, $\ln\gamma_1/x_2$ and $\ln\gamma_2/x_1$ to independent equations and then forced consistent results by adjusting the coefficients of the polynomials. There appears to be no simple way to decide which method of fitting to use even if the choice of a representation equation is already made.

Representation equations appear not to give very quantitative results in a consistency test. However, representation equations are used, later in this thesis, for another purpose. The various forms of the Margules equations which can be conveniently used in a least squares fit are, therefore, developed for later reference.

The Margules four-suffix equations as modified by Carlson and Colburn (7) are

$$\ln\gamma_1 = x_2^2(A + 2x_1(B - A) + x_1(3x_1 - 2)C) \tag{43a}$$

and

$$\ln\gamma_2 = x_1^2(B + 2x_2(A - B) + x_2(1 - 3x_1)C) \tag{43b}$$

The constants A, B and C may be determined by obtaining a least squares fit of Q, Q/x_1x_2, $\ln\gamma_1/\gamma_2$, $\ln\gamma_1$ or $\ln\gamma_2$ data. If the $\ln\gamma_1$ data are used, then it is also possible to use the $\ln\gamma_2$ data to obtain new constants A', B' and C'. If A' = A, B' = B and C' = C the data are consistent. This is not normally the case and some method of making the representation consistent is usually desired. The reason for this desire is not

altogether clear. Apparently the major reason is that it is emotionally more satisfying to use consistent representation equations (even if they are wrong) than to use inconsistent data.

The equation for $\ln\gamma_1/\gamma_2$ as obtained from equations 43a and b is

$$\ln\gamma_1/\gamma_2 = A + 2x_1(B - 2A - C) + 3x_1^2(A - B + 2C) - 4x_1^3C. \quad (44)$$

A least-squares fit of this equation will always yield consistent data. This procedure uses all the data, e.g., γ_1 and γ_2, in the least-squares fit and would seem to be better than fitting equations 43a and b separately. Particularly since the best method of adjusting the two sets of constants is not known.

The equations for Q and Q/x_1x_2 also yield consistent results. These equations are

$$Q = Ax_1 + x_1^2(B - 2A - C) + x_1^3(A - B + 2C) - x_1^4C \quad (45)$$

and

$$Q/x_1x_2 = A + x_1(B - A - C) + x_1^2C. \quad (46)$$

Equations 44, 45 and 46 are second-, third- and fourth-order equations which will later be compared to determine their relative merits in smoothing random uncertainties.

c. Area tests The area tests provide fewer possible approaches than do the slope and representation equation tests. The overall area test and modifications of it which accumulate error from $x_1 = 0$ have been severely criticized in several publications. The symmetrical-area tests suggested by Herrington appear to require excessive effort for the value of the results obtained and also incorporate graphical techniques. Therefore, the only integral test method which remains is the local area test.

The local area test equation may be written as a direct integral of equation 6,

$$f(x) = \sum_{i=1}^{N} \int_a^b x_i d\ln\gamma_i + \int_a^b (\Delta H/RT^2)dT - \int_a^b (\Delta V/RT)dP \qquad (47)$$

or it may be integrated by parts and written as equation 20. In either case it is convenient to adopt the technique of Stevenson and Sater (45) and let the deviation from zero define f(x). The integral limits again indicate that the variable of integration should be evaluated at the adjacent data points designated as a and b.

The major problem associated with the use of equation 47 is the accuracy of a numerical integration technique. Integration formulas which use several points usually require equally spaced data. Vapor-liquid equilibrium data are difficult to obtain at equally spaced intervals in the liquid composition. Therefore, the trapezoid rule is usually used. The use of the trapezoid rule may lead to significant error, however, for activity coefficient curves which are highly unsymmetrical. There is no error involved for perfectly symmetrical curves.

The effect of lack of symmetry in the activity coefficients on the error associated with the use of the trapezoid rule is easily illustrated by using the Margules equations. Use of the trapezoid rule transforms equation 47 to

$$f(a,b) = \sum_{i=1}^{N} \tfrac{1}{2}(x_{ia} + x_{ib})\ln(\gamma_{ib}/\gamma_{ia}) + \int_a^b (\Delta H/RT^2)dT - \int_a^b (\Delta V/RT)dP \qquad (48)$$

The notation f(a,b) indicates the approximate local area test based on the trapezoid rule as opposed to the rigorous formula for f(x). Assume for the moment that $\Delta H = \Delta V = 0$. Then the Margules equations, which

are solutions of the Gibbs-Duhem equation, give f(x) identically zero. Substituting equations 43a and 43b into equation 48 and collecting terms gives

$$f_m(a,b) = -(b - a)^3\{(A - B)/2 - C(a + b - 1)\}. \quad (49)$$

The use of the trapezoid rule with the Margules equations gives results which will not generally be zero. The error is proportional to the cube of the integration interval b - a. The form of $f_m(a,b)$ is particularly simple if the third constant, C, in the Margules equations is zero. The error associated with the trapezoid rule is then seen to be zero if A = B. The choice of A = B represents symmetrical activity coefficient curves for components one and two.

Equation 49 can be used to reduce the error associated with equation 48. Values of A, B and C can be estimated for a system with unsymmetrical activity coefficients and f(x) can be more nearly approximated by

$$f(x) \simeq f(a,b) - f_m(a,b). \quad (50)$$

The fit of the Margules equations to the data need not be particularly good for equation 50 to significantly improve the approximation to f(x). A particularly simple procedure is to take C = 0, extrapolate the activity coefficient curves to $x_1 = 0$ and $x_1 = 1$ where $A = \ln\gamma_1$ and $B = \ln\gamma_2$, respectively, and then to use A and B to calculate values of $f_m(a,b)$ for each interval to be tested. With equal interval data, the error is seen to be constant over the entire composition range. This technique is particularly useful with ternary systems as will be shown by applications to experimental data.

The Margules equations can also be used to illustrate the difference between the local area test and the overall area test. Consider a system with $\Delta H = \Delta V = 0$ with activity coefficient curves that can be described by the two-constant Margules equations. The activity coefficient data for component one are fit by least squares to equation 43a to yield the constants A and B. The activity coefficient data for component two are fit to equation 43b to yield the constants A' and B'. Substituting equations 43a and 43b with the constants A, B and A', B' respectively, into the local area test equation, 47, and integrating gives

$$f(x) = \{x_1^2(B - B' + 2A' - 2A) + (2/3)x_1^3(5A - 5A' - 4B + 4B') + (3/2)x_1^4(B' - B + A' - A)\}_a^b. \quad (51)$$

The overall area test is obtained by taking a = 0 and b = 1 in equation 51. This gives

$$f(1) = (A + B - A' - B')/6. \quad (52)$$

The condition for consistent data is therefore

$$A + B = A' + B'. \quad (53)$$

Using equation 53 to eliminate A and A' in equation 51 results in

$$f(x) = \{3x_1^2(1 - x_1)^2(B - B')\}_a^b. \quad (54)$$

The only condition which will give f(x) identically zero for a series of intervals a, b in equation 54 is B = B'. These comparisons illustrate that the overall area test is satisfied by the simple conditions of equation 53 while the local area test is satisfied only when A = A' and B = B'. The local area test will permit only one relationship between the two activity coefficient curves while the overall area test will permit any number of relationships.

It may also be of interest, although not extremely useful, to note the relationship between the local area test and the slope test. The integration of equation 38 for E_k yields precisely equation 47, f(x), viz.,

$$f(x) = \int_a^b E_k dx_k. \tag{55}$$

Letting k refer to component one and applying the trapezoid rule to equation 55 gives

$$f(a,b) = \bar{E}_1(b - a), \tag{56}$$

where $\bar{E}_1 = (E_{1a} + E_{1b})/2$. Thus, the local area test result is related to the average value of the slope test result multiplied by the scale factor b - a. Thus the average potential error in the activity coefficients could be calculated from equations 42 by using either $f(a,b)/(b - a)$ or $\bar{E}_1$.

The local area test appears to be an excellent thermodynamic consistency test. The application errors associated with the use of the trapezoid rule can be reliably estimated and the test appears to be both a necessary and sufficient condition for the thermodynamic agreement between activity coefficient curves for two components.

2. Calculation of equilibrium values

As was pointed out previously, the method of calculating equilibrium values for one of the experimental variables forces all the random and systematic errors to appear in that variable. There appears to be no convenient way of separating the random and systematic errors in such a procedure. A small difference between the experimental and calculated values may indicate very accurate data. However, if a

large difference results, an investigation for the cause of that error would appear to be quite difficult. The calculation of equilibrium values as a means of testing thermodynamic consistency, therefore, does not yield the quantitative distinction between random and systematic errors that is desired in this work.

3. Discussion

The currently available procedures for testing thermodynamic consistency have been reviewed. The purpose of the review was to identify those consistency tests which could be most accurately applied and which could be used to isolate the composition range in which experimental error occurred.

The tests which were developed primarily for the purpose of calculating equilibrium values are difficult to apply to multicomponent systems. They may also incur cumulative integration errors or require polynomial representations. They appear, therefore, to be of limited value in attempting to indicate the effect of experimental measurement error on phase equilibrium data.

The direct comparison methods are of three basically different types: slope tests; representation equation tests, and; area tests. The slope tests may be applied by either graphical or numerical techniques but in either case the accuracy of slope measurements is low and severely limits the accuracy of the test results. The accurate application of representation equation tests requires either a very versatile representation equation, e.g., orthogonal polynomials, or a prior knowledge of the correct form of the representation equation.

However, the use of orthogonal polynomials may introduce unwarranted inflection points in the representation of the activity coefficients and the correct form of a less versatile representation equation is seldom known. The representation equation tests must therefore be considered as approximate tests, only.

Area tests vary in application from local tests between two adjacent data points to overall tests which may require extrapolation of data. The extended range tests are basically sums or cumulative integrations of the local tests with the overall test being the limit of summation over the entire composition range. These extended range tests have two commonly recognized weaknesses: 1. they may mask large but compensating errors and; 2. small integration errors may cumulate and mask the uncertainty in the data.

The local area test seems to be the most promising test to consider in attempting to distinguish between random and systematic errors. The propagation of error equations are therefore applied to the integral form of the Gibbs-Duhem equation.

B. Application of Propagation of Error Formulas

The effect of random measurement errors on the activity coefficients and the local area consistency test can be predicted by means of the propagation of error formula. This formula states that if a function w is a function of n independent variables z_1, z_2,----, z_n, the variance of w is given by

$$s_w^2 = \sum_{i=1}^{N} \left(\frac{\partial w}{\partial z_i}\right)^2 s_{zi}^2, \tag{57}$$

where s_w^2 is the variance of w and $s_{z_i}^2$ is the variance of z_i. This equation is normally considered to be accurate if the coefficient of variation, viz., s_{z_i}/z_i, is less than 0.2.

The independent variables in this case are those variables that are experimentally measured. They will always include the temperature and pressure and may include either N or N - 1 terms for x_i and y_i. If the composition of component N is calculated from material balances and the N - 1 remaining compositions, then N - 1 terms for x_i and y_i are independent and x_N and y_N are dependent. Alternately, x_N and y_N can be included as independent variables if they are independently measured. In the latter case, the material balance will not necessarily be closed. Letting z be the error in the material balance with the compositions expressed in mole fractions would require

$$z = 1 - \sum_{i=1}^{N} x_i. \tag{58}$$

Then, applying equation 57 gives

$$s_z^2 = \sum_{i=1}^{N} s_{x_i}^2. \tag{59}$$

Let the new values of x_i which close the material balance be calculated as

$$x_i' = z/N + x_i \tag{60}$$

so that $\sum_{i=1}^{N} x_i' = z + \sum_{i=1}^{N} x_i = 1$, by equation 58.

The variance in x_i' is obtained from equation 60 as

$$V_N(x_i') = s_{x_i'}^2 = (1/N)^2 s_z^2 + s_{x_i}^2$$

$$V_N(x_i') = s_{x_i}^2 + (1/N^2)\sum_{i=1}^{N} s_{x_i}^2. \tag{61}$$

where $V_N(x_i')$ represents the variance in x_i for N independently determined compositions. Conversely when only N - 1 compositions are measured, the material balance is closed by calculating x_N as

$$x_N = 1 - \sum_{i=1}^{N-1} x_i \tag{62}$$

so that $V_{N-1}(x_i) = s^2_{x_i}$, i = 1 to N - 1, and

$$V_{N-1}(x_N) = \sum_{i=1}^{N-1} s^2_{x_i} = \sum_{i=1}^{N-1} V_{N-1}(x_i). \tag{63}$$

The notation used here means that the variance $V(x_i)$ is the variance to be used in the propagation of error formula and depends on the method of closing the material balance. The variance $s^2_{x_i}$ is the experimentally determined value for each of the measured compositions.

The effect of choosing N or N - 1 independent variables can be illustrated by considering the relative magnitude of the variances for the two cases. For a binary system the variance of x_i is, by equation 61,

$$V_N(x_i') = s^2_{x_i} + \tfrac{1}{4}(s^2_{x_1} + s^2_{x_2})$$

so that

$$V_N(x_1') = 1.25\ s^2_{x_1} + 0.25\ s^2_{x_2},$$

and

$$V_N(x_2') = 1.25\ s^2_{x_2} + 0.25\ s^2_{x_1}.$$

Equation 63 gives

$$V_{N-1}(x_1) = s^2_{x_1} = V_{N-1}(x_2)$$

Thus if $s^2_{x_1} = s^2_{x_2}$ the result is

$$V_N(x_1') = V_N(x_2') = 1.5\ s^2_{x_1} = 1.5\ V_{N-1}(x_1).$$

Therefore, for a binary system, the smallest variance is always $V_{N-1}(x_i)$. It is obtained by measuring only the one composition which can be most accurately determined and calculating the other from equation 62.

The general result for an N component system with

$s^2_{x_1} = s^2_{x_2} = s^2_{x_3} = \text{----} = s^2_{x_N}$ can be written, by equation 61, as

$$V_N(x_i') = s^2_{x_1} + s^2_{x_1}/N = ((N + 1)/N)s^2_{x_1} .$$

Closing the material balance by the method of equation 60, therefore, always increases the variance by the factor $(N + 1)/N$. Using only $N - 1$ measured compositions gives

$$V_{N-1}(x_i) = s^2_{x_i} = s^2_{x_1}$$

for $i = 1$ to $N - 1$ and by equation 63,

$$V_{N-1}(x_N) = (N - 1)s^2_{x_1} .$$

One measure of comparison for these results is the sum of all the variances as

$$V_N = \sum_{i=1}^{N} V_N(x_i') = \sum_{i=1}^{N} (N + 1/N)s^2_{x_1} = (N + 1)s^2_{x_1}$$

and

$$V_{N-1} = \sum_{i=1}^{N} V_{N-1}(x_i) = \sum_{i=1}^{N-1} s^2_{x_i} + (N - 1)s^2_{x_1} = 2(N - 1)s^2_{x_1} .$$

Then $V_{N-1}/V_N = 2(N - 1)/(N + 1)$ which gives ratios of 2/3, 1, 6/5 and 4/3 for $N = 2$, 3, 4 and 5, respectively. This means that when all the compositions can be measured with equal uncertainty, the smallest sum of variances is obtained by measuring one composition for a binary system and N compositions for N greater than 3. For a ternary system the result is the same by either procedure.

Other methods of comparing the results of these two procedures as well as other methods for distributing the material balance error can of course be developed. Perhaps the material balance error would be more properly distributed by using the equation

$$x_i = z(s^2_{x_i}/s^2_z) + x_i \tag{64}$$

so that equation 61 is replaced by

$$V_N(x_i') = s^2_{x_i} + (s^2_{x_i}/s^2_z)^2 s^2_z = s^2_{x_i}(1 + s^2_{x_i}/s^2_z). \tag{65}$$

The results of closing the material balances by these three different techniques are shown in Table 1. The results seem to indicate that whenever the uncertainty of measuring one composition is greater than the uncertainty in either of the other compositions, only two compositions should be measured. In any event, the propagation of error equations will be developed by assuming that only N - 1 liquid compositions and N - 1 vapor compositions are independent.

In connection with equation 64 it may be worth noting that adjusting the material balance by means of the equation

$$x_i' = z(s_{x_i}/s_z) + x_i$$

is the worst possible procedure even though it may appear to be quite reasonable. Applying the propagation of error formula to this equation yields

$$V_N(x_i) = (s_{x_i}/s_z)^2 s^2_z + s^2_{x_i} = 2s^2_{x_i}.$$

Thus the variance of each component is simply doubled.

1. Variance of f(a,b) and γ_i

The variance of f(a,b) is determined by applying equation 57 to equation 48. Equation 48 contains measured variables evaluated at two points, however, so that the variance of f(a, b) is written

$$s^2_f = s^2_{f_a} + s^2_{f_b}. \tag{66}$$

Then $s^2_{f_a}$ is given by

Table 1. Comparison of three methods for closing material balances

Component Number	$s^2_{x_i}$ x10^6	$V_{N-1}(x_i)$ x10^6 equation 63	$V_N(x'_i)$ x10^6 equation 61	$V_N(x'_i)$ x10^6 equation 65
1	4	4	5.33	5.33
2	4	4	5.33	5.33
3	4	8	5.33	5.33
Sum	12	16	16.00	16.00
1	4	4	5.78	5.00
2	4	4	5.78	5.00
3	8	8	9.78	12.00
Sum	16	16	21.33	22.00
1	4	4	6.67	4.67
2	4	4	6.67	4.67
3	16	8	18.67	21.33
Sum	24	16	32.00	30.67
1	4	4	7.23	4.27
2	9	9	12.23	11.80
3	16	13	19.23	24.83
Sum	29	26	38.69	40.90

$$s^2_{f_a} = \sum_{i=1}^{N-1} \left(\frac{\partial f}{\partial x_{ia}}\right)^2 s^2_{x_{ia}} + \sum_{i=1}^{N-1} \left(\frac{\partial f}{\partial y_{ia}}\right)^2 s^2_{y_{ia}} + \left(\frac{\partial f}{\partial P_a}\right)^2 s^2_{P_a} + \left(\frac{\partial f}{\partial T_a}\right)^2 s^2_{T_a} \qquad (67)$$

with the obvious analogy for $s^2_{f_b}$. The derivatives for the liquid composition are evaluated with x_N taken as a dependent variable so that

$$\frac{\partial f(a,b)}{\partial x_{ia}} = \frac{\partial}{\partial x_{ia}} \left\{ \sum_{j=1}^{N} \tfrac{1}{2}(x_{ja} + x_{jb})(\ln\gamma_{jb} - \ln\gamma_{ja}) \right\}$$

or

$$\partial f(a,b)/\partial x_{ia} = \tfrac{1}{2}\{(\partial x_{ia}/\partial x_{ia})\ln(\gamma_{ib}/\gamma_{ia}) + (\partial x_{Na}/\partial x_{ia})\ln(\gamma_{Nb}/\gamma_{Na})\} -$$
$$\tfrac{1}{2}(x_{ia} + x_{ib})\partial \ln\gamma_{ia}/\partial x_{ia} - \tfrac{1}{2}(x_{Na} + x_{Nb})\partial \ln\gamma_{Na}/\partial x_{ia}. \quad (68)$$

Equation 60 gives

$$\partial x_{Na}/\partial x_{ia} = 1$$

and equation 1, viz.,

$$\gamma_i = y_i P\theta_i / x_i p_i$$

gives

$$\partial \ln\gamma_{ia}/\partial x_{ia} = (1/\gamma_{ia})\partial\gamma_{ia}/\partial x_{ia} = (y_i P/\gamma_{ia} p_i)(-1/x_{ia}^2) = -1/x_{ia}$$

with

$$\partial \ln\gamma_N a/\partial x_{ia} = 1/x_{Na}.$$

Equation 68 then becomes

$$\partial f(a,b)/\partial x_{ia} = \tfrac{1}{2}\{\ln(\gamma_{ib}\gamma_{Na}/\gamma_{ia}\gamma_{Nb}) + (x_{ia} + x_{ib})/x_{ia} -$$
$$(x_{Na} + x_{Nb}/x_{Na}\}. \quad (69)$$

The second term in equation 67 is similarly evaluated and is

$$\partial f(a,b)/\partial y_{ia} = \tfrac{1}{2}\{-(x_{ia} + x_{ib})/y_{ia} + (x_{Na} + x_{Nb})/y_{Na} . \quad (70)$$

The remaining terms are

$$\partial f(a,b)/\partial P_a = -1/P_a + \Delta V_a/RT_a \quad (71)$$

and

$$\partial f(a,b)/\partial T_a = \sum_{i=1}^{N} \tfrac{1}{2}(x_{ia} + x_{ib})d\ln p_{ia}/dT_a - \Delta H_a/RT_a^2. \quad (72)$$

Equation 67 can be more compactly written by letting the expressions in equations 69, 70, 71 and 72 define $K_{x_{ia}}$, $K_{y_{ia}}$, K_{P_a} and K_{T_a} respectively. The equations for $K_{x_{ib}}$, $K_{y_{ib}}$, K_{P_b} and K_{T_a} are then defined as

$$K_{x_{ib}} = \tfrac{1}{2}\{\ln(\gamma_{ib}\gamma_{Na}/\gamma_{ia}\gamma_{Nb}) - (x_{ia} + x_{ib})/x_{ib} + (x_{Na} + x_{Nb})/x_{Nb}\},$$

$$K_{y_{ib}} = \tfrac{1}{2}\{(x_{ia} + x_{ib})/y_{ib} - (x_{Na} + x_{Nb})/y_{Nb}\},$$

$$K_{P_b} = 1/P_b - \Delta V_b/RT_b,$$

and

$$K_{T_b} = \sum_{i=1}^{N} \tfrac{1}{2}(x_{ia} + x_{ib}) d\ln p_{ib}/dT_b + \Delta H_b/RT_b^2.$$

Equation 66 is then written as

$$s_f^2 = \sum_{i=1}^{N-1} (K_{x_{ia}}^2 s_{x_{ia}}^2 + K_{x_{ib}}^2 s_{x_{ib}}^2 + K_{y_{ia}}^2 s_{y_{ia}}^2 + K_{y_{ib}}^2 s_{y_{ib}}^2) + K_{P_a}^2 s_{P_a}^2 + K_{P_b}^2 s_{P_b}^2 + K_{T_a}^2 s_{T_a}^2 + K_{T_b}^2 s_{T_b}^2 . \quad (73)$$

Equation 73 is completely general and applies for any functional relationship for the variances of the independently measured variables. That is the variances may either be constant or some specified function of composition. The variances in x and y depend on the assumption of N - 1 independent variables, however. Equation 73 can be written for N independent variables by deleting x_N, y_N and γ_N from the K_x and K_y terms, increasing the upper limit on the sum from N - 1 to N and appropriately defining the variances of x and y. Note that the variance of the liquid and vapor compositions for component N do not appear in equation 73.

Equation 73 becomes much simpler if the variances for each component are equal and constant over the composition range. Defining

$$s_x^2 = s_{x_{ia}}^2 = s_{x_{ib}}^2,$$

$$s_y^2 = s_{y_{ia}}^2 = s_{y_{ib}}^2,$$

$$s_P^2 = s_{P_a}^2 = s_{P_b}^2,$$

$$s_T^2 = s_{T_a}^2 = s_{T_b}^2$$

and letting

$$K_x^2 = \sum_{i=1}^{N-1} (K_{x_{ia}}^2 + K_{x_{ib}}^2),$$

with similar expressions for K_y, K_P and K_T, equation 73 is written

$$s_f^2 = K_x^2 s_x^2 + K_y^2 s_y^2 + K_P^2 s_P^2 + K_T^2 s_T^2. \tag{74}$$

This equation for the variance of f(a,b) will probably be the most useful one in analyzing the majority of the currently available vapor-liquid equilibrium data.

One of the major goals of this work is to relate the effect of random measurement errors in each of the measured variables to the values of f(a,b) to be expected from the consistency test. This can be accomplished by applying the propagation of error formula to the definition of the activity coefficient. Applying equation 57 to equation 1 gives

$$s_{\gamma_i}^2/\gamma_i^2 = s_{x_i}^2/x_i^2 + s_{y_i}^2/y_i^2 + s_P^2/P^2 + (d\ln p_i/dT)^2 s_T^2. \tag{75}$$

The last term in equation 75, for example, is obtained from

$$\begin{aligned} \partial\gamma_i/\partial T &= (\partial\gamma_i/\partial p_i)/(dp_i/dT) \\ &= (y_i\theta_i P/x_i)(-1/p_i)^2 dp_i/dT \\ &= (\gamma_i/p_i)dp_i/dT \end{aligned}$$

so that

$$(\partial\gamma_i/\partial T)^2 = \gamma_i^2 (d\ln p_i/dT)^2.$$

The factor γ_i^2 arises in each term and is transposed to the left side of the equation so that each of the terms in equation 75 is expressed as the square of the coefficient of variation.

2. Statistical meaning of s_f and s_γ

The statistical concepts found most convenient for relating f(a,b)

and errors in γ_i to their variances are briefly stated for future reference. Normally distributed random measurement errors are assumed.

Consider an experimental technique for obtaining vapor-liquid equilibrium data which incurs no errors except the random errors associated with the experimentally measured variables. Assume that the variance in each of these variables is precisely known. If a very large number, M, of data points is obtained, the values of f(a,b) will be randomly distributed about zero. For M approaching infinity, exactly half of the f(a,b) values will be positive and half will be negative. The distribution of these f(a,b) values will also be described by the normal curve of error. Therefore, 68.3 per cent of the f(a,b) will be smaller in absolute value than s_f. Similarly 50 per cent will be smaller than 0.67449 s_f and 95.5 per cent will be smaller than $2s_f$. The $\pm s_f$, $\pm 0.67449 s_f$ and the $\pm 2s_f$ bounds define confidence limits which are referred to here as the 68 per cent, 50 per cent and 95 per cent confidence regions, respectively.

When a finite number of data points are taken, this is equivalent to taking a finite sample of values from an infinite population. The probability of obtaining a given fraction of these points within a specified confidence limit can be calculated from the binomial expansion. Let the per cent of these f(a,b) values which lies within a given confidence limit be defined as p_f. The variance of p_f is then given, to a good approximation, by

$$s^2_{p_f} = q(100 - q)/M, \tag{76}$$

where q = 50 for the 50 per cent confidence region and 68.3 for the 68 per cent confidence region. Therefore, the two standard deviation

reliability on p_f is

$$p_f = q \pm 2\sqrt{q(100 - q)/M}. \quad (77)$$

the two standard deviation reliability is the usual criterion for testing a set of data and corresponds approximately to a probability of 0.05. That is, the chances are less than 1 in 20 of selecting a sample of M data points for which p_f is less than $q - 2s_{p_f}$ or greater than $q + 2s_{p_f}$. The values of p_f which have a greater than 0.05 probability of occurring are enclosed within the bounds of the curves shown in Figure 3. The curves in Figure 3 show both values of q, i.e., 50 per cent and 68.3 per cent. Both values were used in this work and the choice between them seems to be largely one of personal preference. The 50 per cent confidence region may have slightly more statistical significance but the 68 per cent confidence region is more easily calculated.

The curves in Figure 3 are used to test the reliability of a set of vapor-liquid equilibrium data. The consistency test is applied to a set of data for which the variance of each of the independently measured variables is known. The values of f(a,b) and s_f are calculated for each interval between adjacent data points. The value of p_f is then determined. If the value of p_f lies within the appropriate curves in Figure 3, the data should not be rejected as unreliable. If the value of p_f lies outside the curves in Figure 3, the probability is less than 1 in 20 that the data are a sample from the infinite population of points described by the variances associated with the measured variables. Three possibilities must then be considered: 1. this

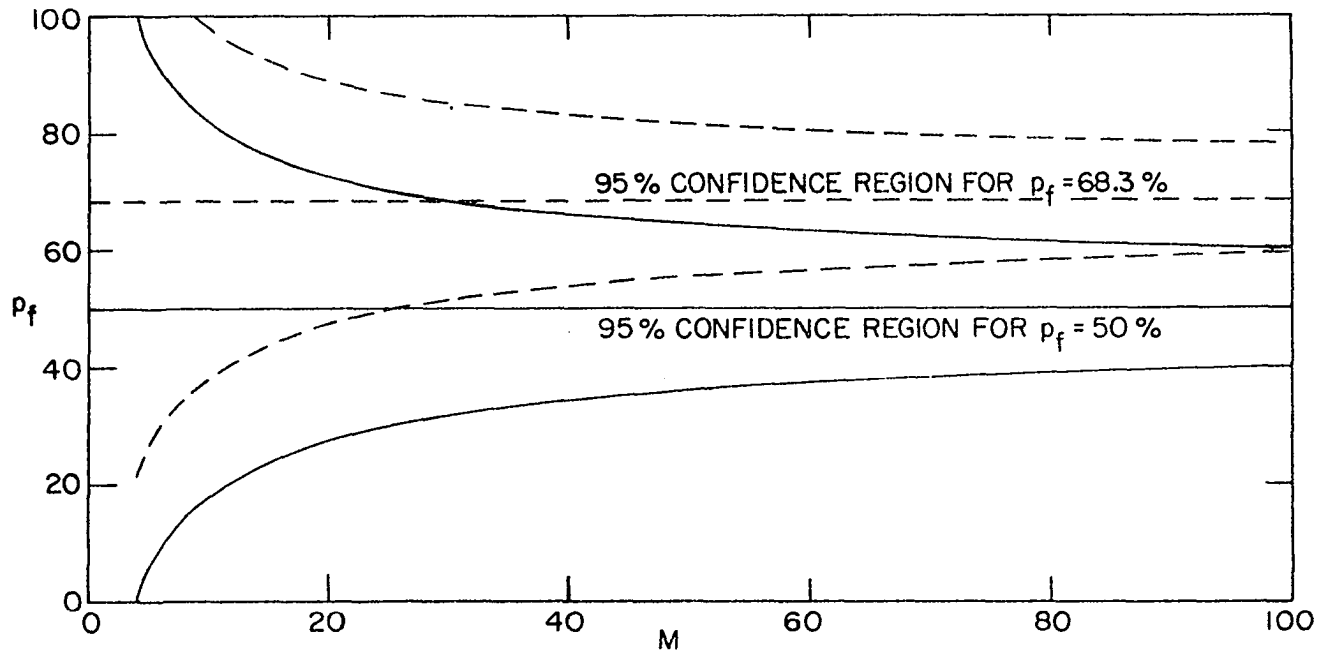

Figure 3. The 95 per cent confidence regions for p_f

set of data happens to be the 1 in 20 set and the data are accurately measured; 2. the variances used are incorrect; or 3. the data contain systematic error in addition to the random error. The first possibility must be left to the discretion of the individual testing the data. Several methods are available to aid in distinguishing between the last two possibilities.

Systematic errors and errors caused by neglecting pressure corrections, ΔH or ΔV will cause nonrandom trends in f(a,b). If these errors are of significant magnitude their presence will usually be obvious from an inspection of the f(a,b) values. The statistical significance of such trends can be checked by applying tests for randomness. Based on the limited number of cases in this work, the value of these tests is low unless M is on the order of 25 or 30. One fairly simple check, the chi-square test, is presented for checking the distribution about zero. Letting N_+ be the number of positive values of f(a,b), the value of chi-square is

$$\chi^2 = 4(N_+ - M/2)^2/M. \tag{78}$$

The probability is less than 1 in 20 of obtaining a chi-square greater than 3.84. For a set of 10 values of f(a,b) this test would not reject a distribution of 8 positive and 2 negative values for which chi-square is 3.6. A distribution of 9 to 1 would be rejected with chi-square equal to 6.4. Statistical tests are thus not very discerning with respect to a gross displacement of the data points and a critical inspection of the distribution may be more meaningful. For example, if most of the points are positive and most of the points which lie outside the confidence region are also positive, the possibilities which

could cause a displacement of the data should be investigated.

The statistical techniques described above also apply to the relationship between the error in γ_i and s_{γ_i}. The error in the activity coefficient can not be determined from experimental data alone, however, because the expected value is not normally known. Expressions for the error in the activity coefficient can be obtained only when the random error is simulated or when the activity coefficients are in some way smoothed. This subject is discussed in Section III, D.

C. Characterization of the Consistency Test

The effect of measurement errors in each of the independent variables x, y, P and T is illustrated by simulating vapor-liquid equilibrium data and calculating s_f. Hypothetical systems are chosen to illustrate the behavior of positive and negative deviations from Raoult's law, symmetrical and unsymmetrical activity coefficient curves, and azeotropes. In all cases the systems are described by the three-suffix Margules equations (equations 43 and C = 0). The vapor pressure equation for the constant pressure systems is $\ln p_i = D_i + E_i/T$ with T in ^{0}K. Tables 2 and 3 list the pertinent characteristics of the systems. Figure 4 shows the activity coefficient curves for these four systems. Figure 5 shows the x - y relationships. Examples of systems which exhibit characteristics similar to systems I, II, III and IV are, respectively, diisopropyl ether-isopropyl alcohol, water-n-butyl alcohol, ethanol-chloroform, and acetone-chloroform (7).

The general procedure followed in simulating binary test systems was to generate a set of M error free data points and then to add

Table 2. Characteristics of the simulated constant pressure systems

System	A	B	P mmHq	D_1	E_1	D_2	E_2
I	1.0	1.0	760	17	-3600	22	-5400
II	2.0	1.0	760	28	-8000	22	-6000

Table 3. Characteristics of the simulated constant temperature systems

System	A	B	T ^0K	p1 mmHq	p_2 mmHq	dp_1/dT mmHq/^0K	dp_2/dT mmHq/^0K
III	1.5	0.5	318	170	430	8	15
IV	-1.0	-0.8	308	350	300	12	10

normally distributed random error to the desired variable. This provided a means for checking the effect on the activity coefficients as well as testing the accuracy of the consistency test equations. The error free points were obtained by selecting a value of M and thus M - 1 equally spaced intervals in x_1 given by $x_{1b} - x_{1a} = 1/(M + 1)$. The activity coefficients were calculated from the Margules equations for each value of x_1. For constant temperature systems the total pressure was then calculated from the equation

$$P = x_1\gamma_1P_1 + x_2\gamma_2P_2$$

by using the vapor pressure data in Table 3. For constant pressure systems the vapor pressure equations, with the constants from Table 2,

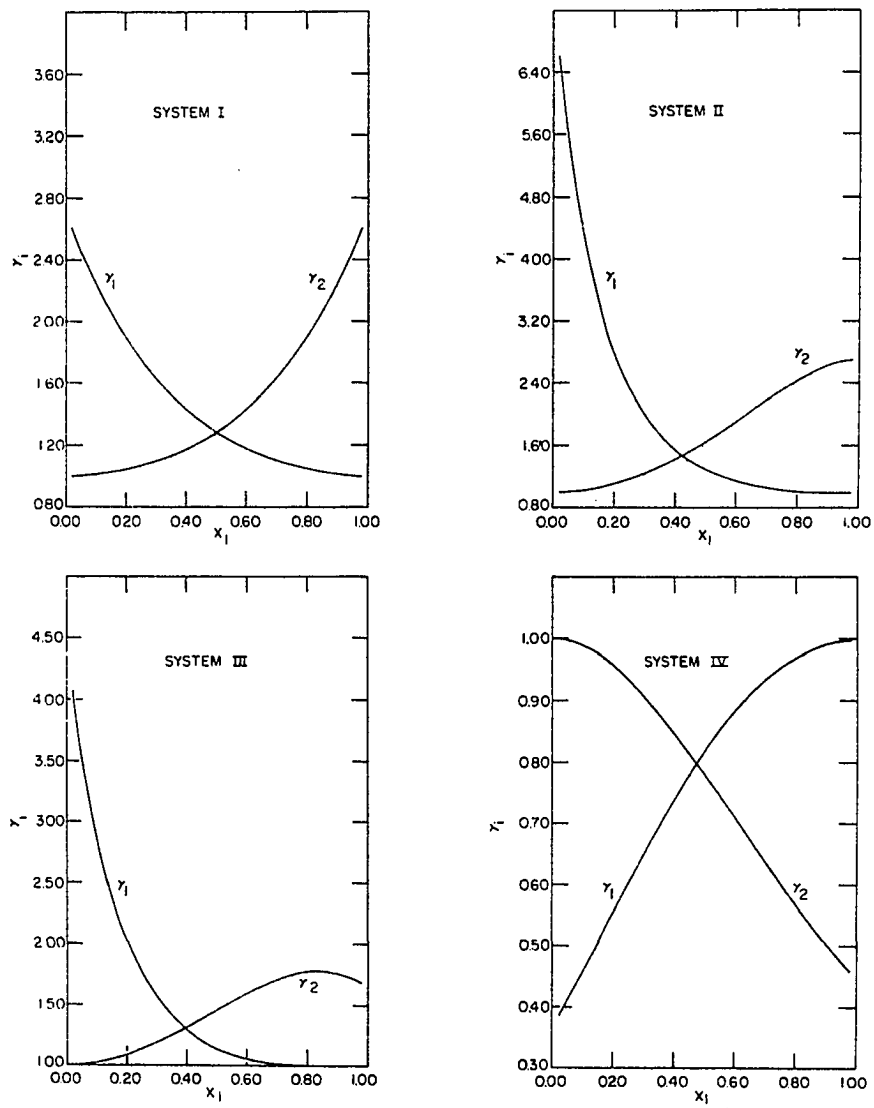

Figure 4. The activity coefficient curves for the four hypothetical systems

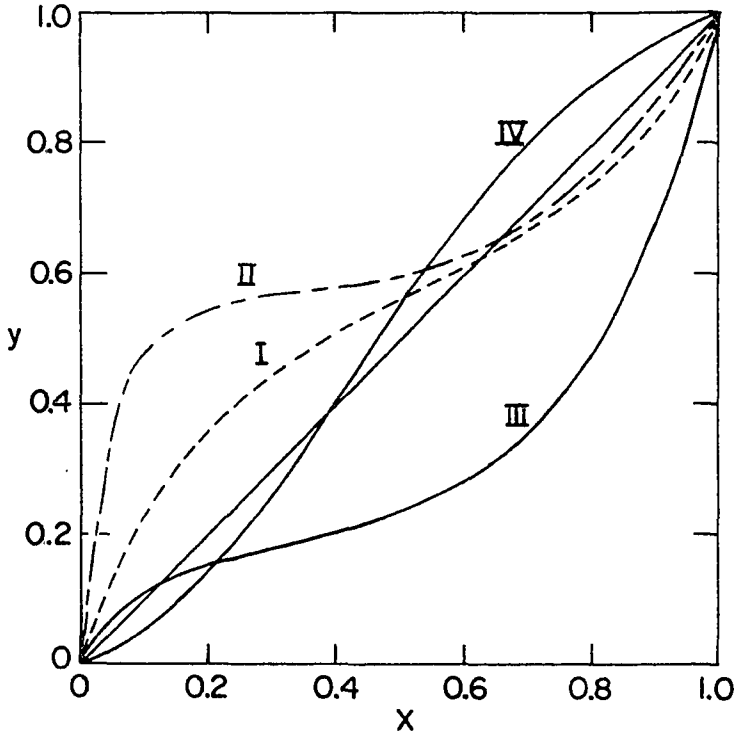

Figure 5. The x-y relationship for the four hypothetical systems

were used to calculate the temperature by iteration. The vapor composition was then calculated from the equation $y_1 = x_1\gamma_1 P_1/P$.

The random error in each variable was generated by using IBM subroutine Gauss. The standard deviations s_x, s_y, s_p and s_T were specified as input to the subroutine and independent random errors of zero mean were calculated for each of the four variables at each of the M points. A new set of data which contained imposed random error was obtained by adding the calculated error terms to each of the four variables. The system containing imposed random error was distinguished by a superscript e as x_1^e, y_1^e, P^e and T^e. Activity coefficients, γ_1^e, and γ_2^e, were then calculated in the usual manner. Consistency tests were applied to this data to check the validity of the assumptions made in deriving the equations for the variance of f and to illustrate the characteristics of the consistency test. Checks for systematic error were not required because only normally distributed random error were imposed on the previously consistent system.

1. Verification of consistency test procedure

The statistical tests suggested above as a means of interpreting consistency test results, require the assumption of a normal distribution in f. The random error in functions which are products, ratios or logarithms of variables containing normally distributed random error are not necessarily normally distributed. The quantity f(a,b) is obtained from products, ratios and logarithms of the variables x^e, y^e, P^e and T^e. The validity of assuming a normal distribution

is therefore examined. The results are independent of the system used and only the results from system III, Table 3, are discussed.

The procedure used was to pick a value of M, e.g., 19, impose random error on each variable, calculate f(a,b) and s_f for each interval in x_1 and then repeat the procedure a sufficient number of times, e.g., 200, to obtain a reliable distribution in f(a,b) for each interval. The values of s_x, s_y, s_P and s_T remained constant but new random errors were imposed in each iteration by selecting a different sample from the infinite population. The value of s_f was the same for each iteration because s_x, s_y, s_P and s_T were held constant. If the f(a,b) values were normally distributed, 68.3 per cent of them would lie within $\pm s_f$ with a variance given by equation 77.

The results of applying the procedure twice (M = 19 with 200 iterations) are shown in Table 4 for two selected intervals of x_1. The levels of error indicated in the first column are approximately midpoints of the range of error studied. The expression

$$\{\Sigma f(a,b)^2/199\}^{\frac{1}{2}}$$

is the formula for calculating the standard deviation of f and should approximate the value of s_f obtained from equation 74. From equation 77 $p_f = 68.3 \pm 6.5$ so that the chances are less than 1 in 20 of obtaining p_f greater than 74.8 or less than 61.8 if the distribution is normal. The three values of p_f indicated by asterisks in Table 4 are outside this range. Further examination of the distribution of f(a,b) shows that the greatest deviation from the normal (or Gaussian) distribution occurs when only s_y is nonzero. The deviation varies with the choice of the interval, $x_{ib} - x_{ia}$. Thus, as shown in Table 4 for

Table 4. Test for normal distribution in f

Imposed error	$\frac{x_{1b} - x_{1a}}{\{\frac{\Sigma f(a,b)^2}{199}\}^{\frac{1}{2}}} = \frac{0.10 - 0.05}{s_f}$		p_f	$\frac{x_{1b} - x_{1a}}{\{\frac{\Sigma f(a,b)^2}{199}\}^{\frac{1}{2}}} = \frac{0.55 - 0.50}{s_f}$		p_f
s_x=0.004,	0.00162	0.00177	73.5	0.00030	0.00030	65.0
$s_y=s_p=s_T$=0	0.00184	0.00177	67.0	0.00032	0.00030	62.5
s_y=0.004,	0.00155	0.00200	81.5[a]	0.00891	0.00836	63.5
$s_x=x_p=x_T$=0	0.00155	0.00200	81.0[a]	0.00899	0.00836	67.0
s_p=2.0,	0.00620	0.00630	72.0	0.00650	0.00685	73.5
$s_x=s_y=s_T$=0	0.00643	0.00630	71.5	0.00632	0.00685	75.0[a]
s_T=0.2,	0.01028	0.01066	68.0	0.01312	0.01241	67.0
$s_x=s_y=s_p$=0	0.01116	0.01066	66.5	0.01211	0.01241	69.5
$s_x=s_y$=0.004,	0.01268	0.01267	71.0	0.01660	0.01645	69.0
s_p=2.0, s_T=0.2	0.01264	0.01267	67.0	0.01706	0.01645	66.0

[a]The probability is less than 1 in 20 of obtaining numbers this large.

s_y = 0.004, the distribution is significantly different from normal when $x_{ib} - x_{ia}$ = 0.10 - 0.05 but satisfactory results are obtained for $x_{ib} - x_{ia}$ = 0.55 - 0.50. Fortunately, the average result over all 18 intervals of width 0.05 indicates that the distribution of f(a,b) is sufficiently near normal to permit use of p_f values calculated over the composition range.

The test procedure suggested here requires that the relationship of one value for f(a,b) in each interval of width 1/(M + 1) be compared to the confidence region for that same interval. The value of p_f is then obtained by summing the results over all intervals. If the distribution in any of these intervals is not Gaussian, the fiducial

limits given by equation 77 may again be in error. The 200 random error samples discussed above were also used to calculate 200 values of p_f by summing the number of f(a,b) values which were inside their respective confidence region and dividing by 18. The average values of p_f and the estimated standard deviation for p_f, viz., $\{\Sigma(p_f - 68.3)^2/199\}^{\frac{1}{2}}$ are presented in Table 5. The value of s_{p_f} calculated from equation 76 is also shown. The average value for p_f of 67.5 is low primarily because of the low values obtained for $s_y = 0.004$. The average value of the remaining 8 numbers is 68.1. Since actual data will normally contain error in all four variables the average values of p_f are considered satisfactory. The calculated standard deviations, however, are uniformly high by an average of 1.6. Similar results are obtained by applying the same procedure to M = 9 and 39 points. For M = 9, the calculated standard deviation is 19.0 as compared to 16.5 obtained from equation 76. For M = 39, the results are 8.4 and 7.5 respectively. These results are shown in Figure 6.

The results in Figure 6 and Table 5 indicate that some deviation from a normal distribution exists. This deviation appears to be caused primarily by the existence of slightly larger tails on the distribution curve than would be expected for a Gaussion distribution. The existence of a few large positive and a few large negative values of f(a,b) causes the high values for the calculated standard deviation but does not significantly affect the average value of p_f. The difference between the calculated and predicted standard deviations is small enough, however, that a normal distribution in f can be assumed without

Table 5. Average value of p_f and the standard deviation of p_f

Imposed error	Average p_f per cent	$\{\frac{\Sigma(p_f - 68.3)^2}{199}\}^{\frac{1}{2}}$[a]
s_x=0.004, s_y=s_p=s_T=0	68.0	12.5
	66.3	12.5
s_y=0.004, s_x=s_p=s_T=0	65.3	12.3
	64.8	13.8
s_p=2.0, s_x=s_y=s_T=0	68.2	13.2
	67.9	11.7
s_T=0.20, s_x=s_y=s_p=0	68.2	12.5
	68.3	11.7
s_x=s_y=0.004, s_p=2, s_T=0.2	69.6	12.7
	68.4	12.8
Average	67.5	12.6

[a]From equation 76: $s_{p_f} = \{68.3(31.7)/18\}^{\frac{1}{2}} = 11.0$.

incurring serious error. For example, for M = 19 (18 values of f(a,b)) the lower limit for p_f is 68.3 - 2(12.6) = 43.1 from Table 5 or 68.3 - 2(11.0) = 46.3 for equation 76. Then, with 8 of the 18 values inside the 68 per cent confidence region p_f is 44.4 per cent which is acceptable by one criterion and not by the other. However, with only 7 of the 18 values inside the confidence region, p_f is 38.9 per cent and unacceptable by either criterion. In no case will the difference between the two standard deviations permit a difference of more than one point inside or outside of the confidence region. The fiducial limits obtained from equation 77 and plotted in Figure 3 should, therefore, be used.

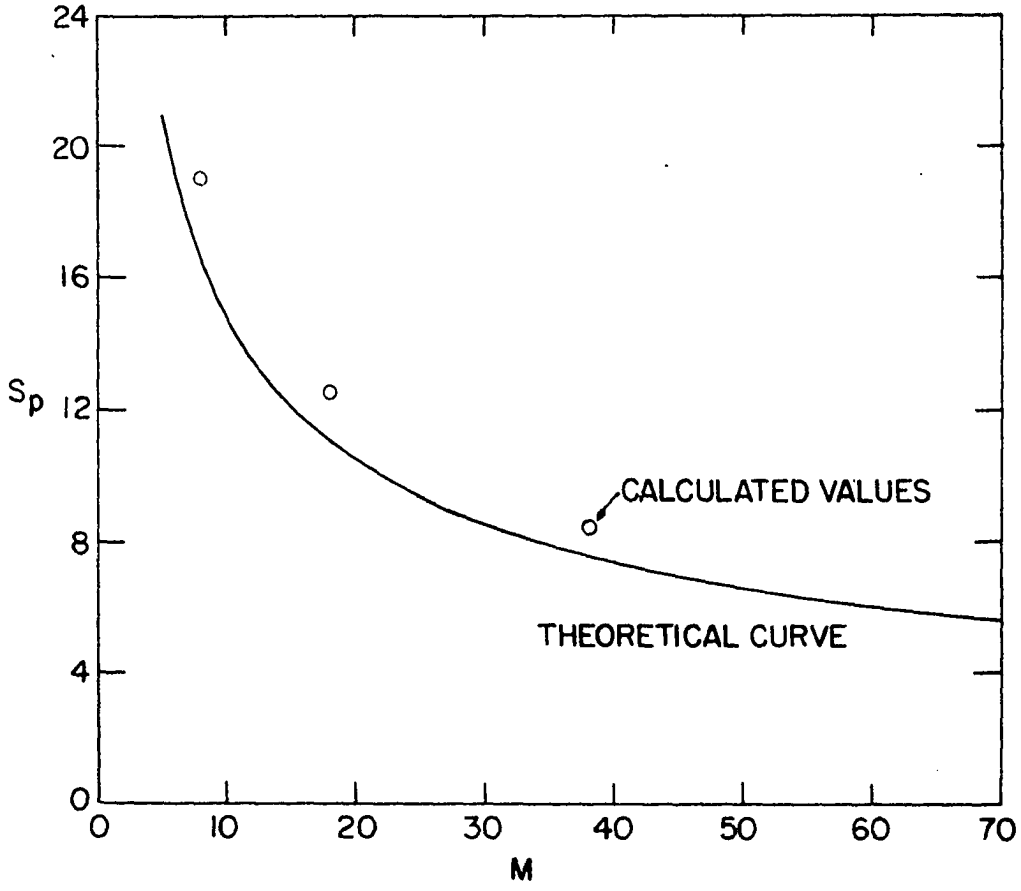

Figure 6. Calculated and theoretical standard deviations for p_f

A ternary system was also simulated and the consistency test procedure verified. No significant differences from the binary system results were noted. Ternary systems will, therefore, be considered only in testing actual data.

2. Effect of random error on s_f

The hypothetical systems presented in Tables 2 and 3 are used to illustrate the effect of random measurement errors on the consistency test. Equation 74, which requires the variances in x, y, P and T to be independent of x, is used to calculate s_f. Only one level of error in each variable is considered because the coefficient terms, e.g., K_x^2, do not depend on the level of error. The values of the coefficient terms are fixed by specifying the chemical system and therefore four curves which show the effect of error in each variable independently are sufficient to determine the effect of any other combination of errors.

Average magnitudes for s_x, s_y, s_P and s_T are difficult to establish because of the variety of techniques used to obtain vapor-liquid equilibrium data and because of the variety of chemical and physical conditions possible. The values of $s_x = 0.004$, $s_y = 0.004$, $s_P = 2.0$ and $s_T = 0.2$ are used because they are convenient for presentation. The results for s_f are shown as 68 per cent confidence regions in Figure 7 for the four hypothetical systems.

These results are based on 48 equal intervals of width $x_{ib} - x_{ia} = 0.02$. The results are calculated from the set of data containing error, i.e., x_1^e, y_1^e, P^e and T^e, and therefore show some fluctuation in the curves.

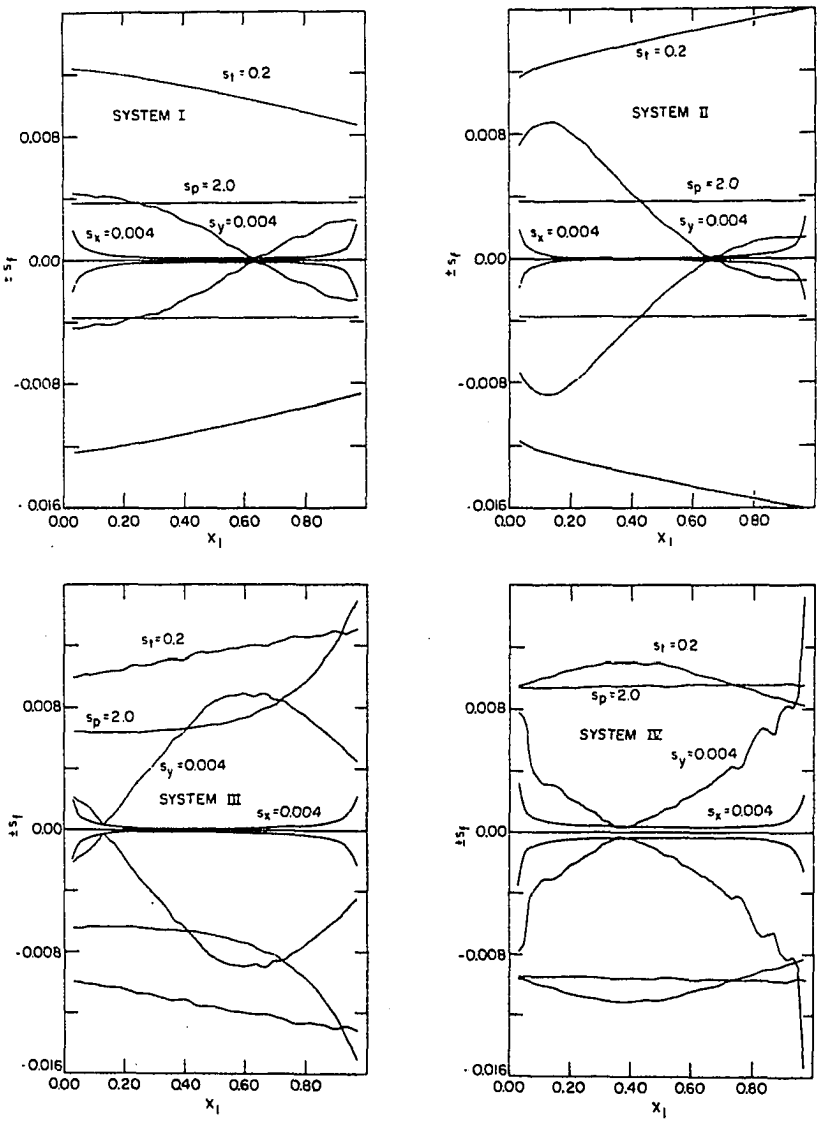

Figure 7. The confidence regions for the local area test with measurement error in only one variable

The results in Figure 7, e.g., system I, indicate that the effect of random error in x_1 is significantly less than the effect of error in the other variables. This means that if the measuring techniques for x and y have equal uncertainties, the consistency test will be significantly affected only by the uncertainty in y for x_1 between 0.1 and 0.9. The uncertainty in x becomes significant only for x_1 between 0.0 and 0.1 or between 0.9 and 1.0. This illustrates that a fixed average value for f(a,b) can not be used to decide whether a set of data are accurately measured. Some estimate of the uncertainty in the measured variables is required to accurately interpret the results of a consistency test.

The hour glass shape of the curve for $s_y = 0.004$ in Figure 7 (system I) is caused by the existence of an azeotrope at $x_1 = 0.63$. Equation 70 for $K_{y_{ia}}$, written for a binary system, involves the difference between x_1/y_1 and x_2/y_2. Near an azeotrope the values of x_1/y_1 and x_2/y_2 both approach 1 and the difference approaches zero. Therefore, the consistency test is not significantly affected by uncertainties in measuring y near an azeotrope. From an analysis of experimental data which contained significant uncertainty only in y_1, the values of f(a,b) might indicate very accurate data near the azeotrope when the opposite may actually be true. This again illustrates the importance of including an error analysis with a consistency test.

The curves for the confidence region for error in pressure and temperature are nearly linear. For a constant pressure system with only s_P nonzero, the value of s_f should be a constant as shown in Figure 7 for systems I and II. If s_P is nonzero, s_f varies as 1/P at constant temperature.

When only s_T is nonzero, equation 72 indicates that s_f is a linear combination in x_1 of $d\ln p_1/dT$ and $d\ln p_2/dT$. The small temperature change for system I adds a slight curvature to the curve shown in Figure 7.

Figure 7 shows similar results for all four hypothetical systems. All of the systems contain azeotropes and the curves for $s_y = 0.004$ are significantly different for each system. Although the curves for $s_x = 0.004$ reflect the unsymmetric nature of a system, all the curves except for the $s_y = 0.004$ curve are remarkably independent of the system. The primary differences among systems are caused by the pressure level and the nature of the vapor pressure versus temperature relationships.

It would be quite useful for someone planning to obtain experimental data to be able to predict the relative effect of experimental measurement uncertainties. The following equation is therefore offered as a rough estimate,

$$s_f^2 = (0.1s_x)^2 + (0.7s_y)^2 + (1.4s_p/P)^2 + (0.7s_T)^2\{(d\ln p_1/dT)^2 + (d\ln p_2/dT)^2\}. \quad (79)$$

Equation 79 gives a reasonable average value for the effect of s_x for x_1 between 0.1 and 0.9. The effect of s_y is extremely difficult to predict but equation 79 gives reasonable averages for the hypothetical systems studied here. The correct factor for the effect of s_p is $\sqrt{2}/P$ for a constant pressure system so that equation 79 is as accurate as the choice of the average pressure. Similarly the effect of temperature is accurately predicted by equation 79 for a constant temperature

system at $x_1 = 0.5$.

The effect of combined errors in all four variables, $s_x = 0.004$, $s_y = 0.004$, $s_P = 2.0$ and $s_T = 0.2$, on the standard deviation, s_f, is shown in Figure 8 for each of the hypothetical systems. The curves are quite similar and indicate that the size of the confidence region depends more on the level of error than on the nature of the system.

3. Estimating ΔH and ΔV from consistency tests

Sets of vapor-liquid equilibrium data are frequently published which give nonrandom trends in f(a,b). This nonrandom trend may be caused by neglecting temperature and pressure effects in the consistency test or it may be caused by errors in the experimental procedure. A procedure for detecting which effect is present would be desirable.

Values for θ_i can usually be estimated with reasonable accuracy from equations of state or generalized procedures. If θ_i has been neglected in the test procedure, the best course seems to be to estimate θ_i, correct the activity coefficients and test the data again. Numerical values for ΔH and ΔV are not easily estimated. Therefore, if neglecting the heat of mixing or volume change of mixing is suspected of contributing to nonrandom trends in f(a,b), a different approach might be helpful.

The ΔH and ΔV functions for binary systems depend on x_1 and x_2 in much the same manner as Q. They can often be represented by polynomials similar in form to those proposed by Redlich and Kister (41) and presented here as equation 17, viz.,

$$\Delta H = x_1(1 - x_1)\{A' + B'(x_2 - x_1)\}. \qquad (17)$$

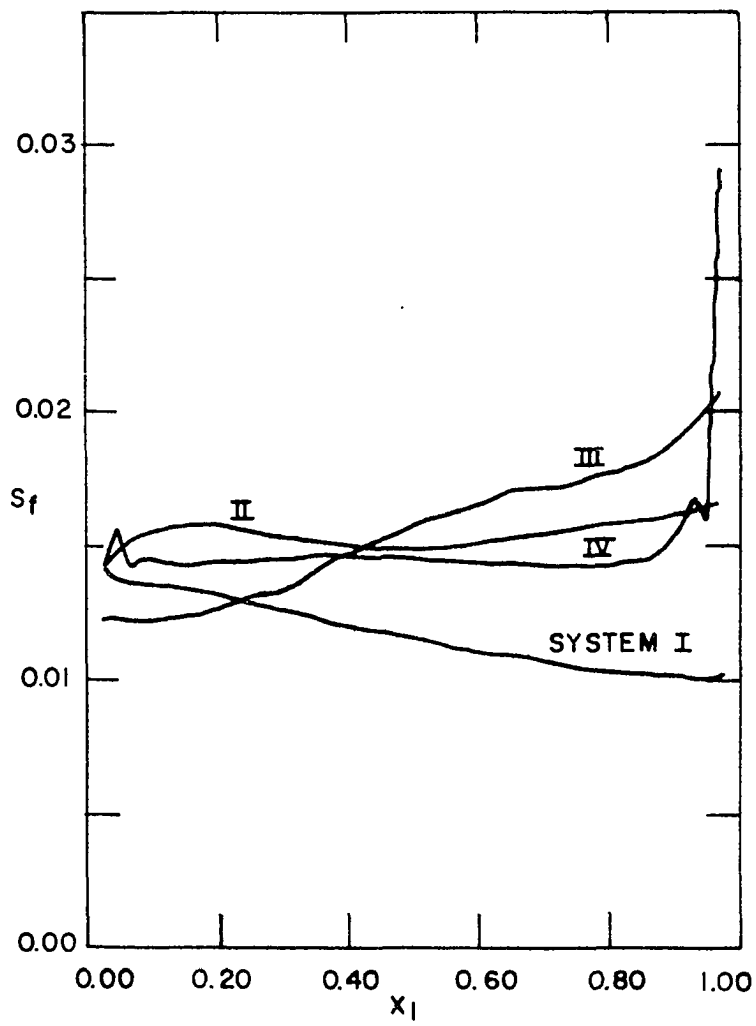

Figure 8. The s_f curves for the four hypothetical systems with measurement error in all four variables

Since the contribution of random error to f(a,b) is tested by means of the propagation of error equations, all systematic error should be removed from f(a,b) and attributed to other types of error. One method for removing the systematic error from f(a,b) would be to assume that ΔH or ΔV is described by a polynomial such as equation 17 and then estimate values of A' and B' which remove the nonrandom trend.

Equation 48 for f(a,b) can be written as

$$f(a,b) = \sum_{i=1}^{N} \tfrac{1}{2}(x_{ia} + x_{ib})\ln(\gamma_{ib}/\gamma_{ia}) + (\overline{\Delta H/RT^2})(T_b - T_a) - (\overline{\Delta V/RT})(P_b - P_a) \qquad (80)$$

by using the trapezoid rule and average values of the quantities in parenthesis to evaluate the heat of mixing and volume change of mixing integrals. The heat of mixing term in equation 80 could also be evaluated as $(\overline{\Delta H}/R)(1/T_a - 1/T_b)$ but this does not appear to provide any real improvement in the present procedure. Let the first term on the right side of equation 80 be designated $f_{T,P}$ to indicate the test results obtained by assuming constant temperature and pressure. Equation 80 is then written as

$$f(a,b) = f_{T,P} + (\overline{\Delta H/RT^2})(T_b - T_a) - (\overline{\Delta V/RT})(P_b - P_a). \qquad (81)$$

When a set of constant pressure data is tested by neglecting the heat of mixing, for example, then $f_{T,P}$ is determined. The results for $f_{T,P}(a,b)$ will always contain some random error but should contain nonrandom error only as a result of neglecting the heat of mixing or volume change of mixing. The results for f(a,b), however, should contain systematic error only if the experimental procedure was incorrect. The suggested test procedure is to assume that f(a,b) should contain

only random error. Therefore, the nonrandom error should be totally attributed to the effect of ΔH or ΔV as

$$f_{T,P} = -(\overline{\Delta H/RT^2})(T_b - T_a) + (\overline{\Delta V/RT})(P_b - P_a).$$

The ΔH and ΔV terms are then estimated from

$$(\overline{\Delta H/RT^2}) = -f_{T,P}/(T_b - T_a) \tag{82}$$

or

$$(\overline{\Delta V/RT}) = f_{T,P}/(P_b - P_a). \tag{83}$$

The consistency test results according to equation 82 or 83 are then plotted against x_1 and smoothed to indicate the magnitude of the ΔH or ΔV terms that would be required to consider the data accurately measured. An estimate of the accuracy of the experimental procedure can be made if some means of estimating ΔH or ΔV is available. An estimate of ΔH or ΔV might be made if the system being tested is a member of an homologous series. Alternately the value of ΔH or ΔV at one value of x_1 might be known.

A polynomial such as equation 17 and a least-squares fit can be used in place of the graphical technique suggested above. The major difficulty in using a polynomial is in choosing the proper form of the polynomial before applying the regression analysis. The advantage of the polynomial and regression analysis procedure is that less bias may be introduced in interpreting whether a given distribution is random. The best procedure will probably be to use both the graphical and regression analysis methods for most cases.

Both of these procedures are necessarily approximate. A given distribution of values for $f(a,b)$ or $f_{T,P}$ can be declared nonrandom only within some given probability. Therefore, the results of applying

equation 82 or 83 should be used only to aid in assessing the accuracy of the vapor-liquid equilibrium data. The estimated values of ΔH and ΔV will always contain random uncertainty and will often contain systematic error arising from the experimental procedure. The estimate of ΔH and ΔV may contain significantly larger random errors than $f_{T,P}$ because of the division by ΔT or ΔP in equations 82 or 83. If the uncertainty of measuring temperature, for example, is greater than 10 per cent of the difference $T_b - T_a$, then the uncertainty of $T_b - T_a$ may be greater than 20 per cent and ΔH will be very difficult to estimate. Confidence regions could be developed for the estimate of ΔH but they would appear to be of little value in most cases.

The regression analysis procedure is illustrated here by using equation 17 and system III of Table 3. Equation 17 is written as

$$\Delta V/RT = x_1 x_2 \{A' + B'(x_2 - x_1)\}. \tag{84}$$

Values of $A' = 0.006$ and $B' = 0.002$ are chosen and the $\Delta V/RT$ curve shown in Figure 9 is generated. Equations 84 and 80 are then used to evaluate f(a,b). The calculated values of f(a,b) are assumed to be $f_{T,P}$ and equations 83 and 84 are used to determine new values of A' and B'. Using system III with $M = 9$, $s_x = 0.002$, $s_y = 0.002$, $s_P = 1.0$ and $s_T = 0.1$, the values of $A' = 0.0073$ and $B' = -0.0047$ were obtained by a least-squares procedure. The curve obtained with these constants is also shown in Figure 9. The values of $f_{T,P}$ and the 68 per cent confidence region are shown in Figure 10. Also shown are the values of f(a,b) obtained by correcting for the volume change of mixing as estimated from the regression analysis. Only two of the $f_{T,P}$ points ($p_f = 25\%$) are inside the confidence region while 75 per cent of the

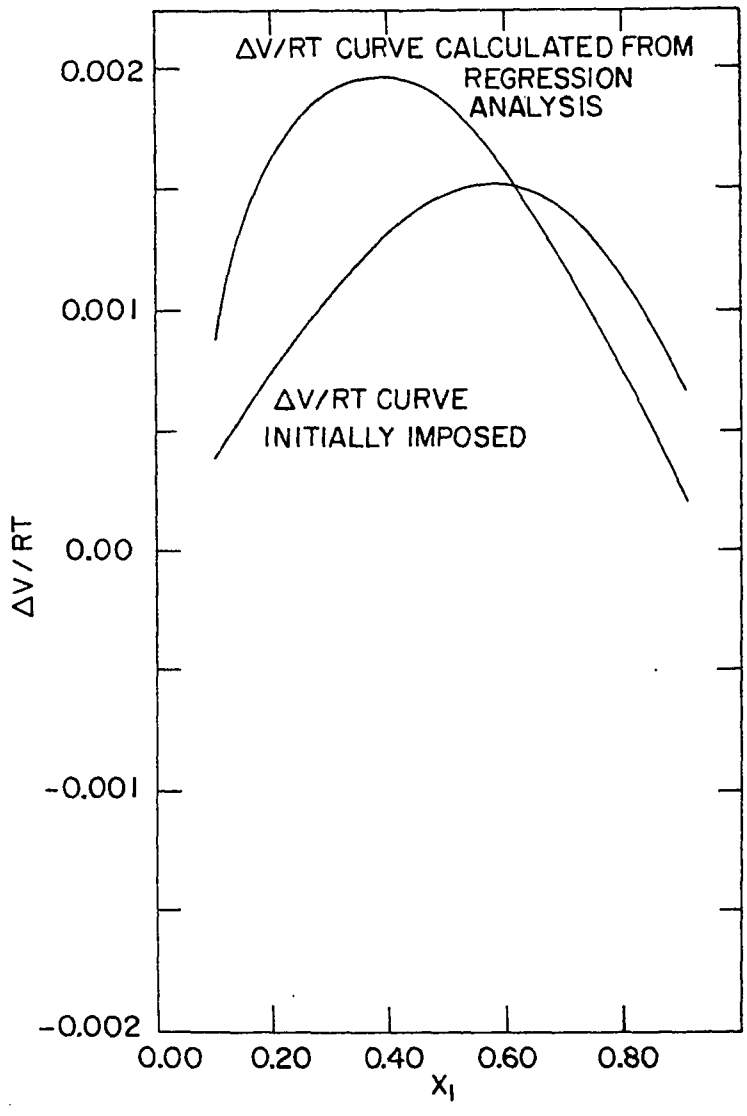

Figure 9. Imposed and calculated curves for ΔV/RT

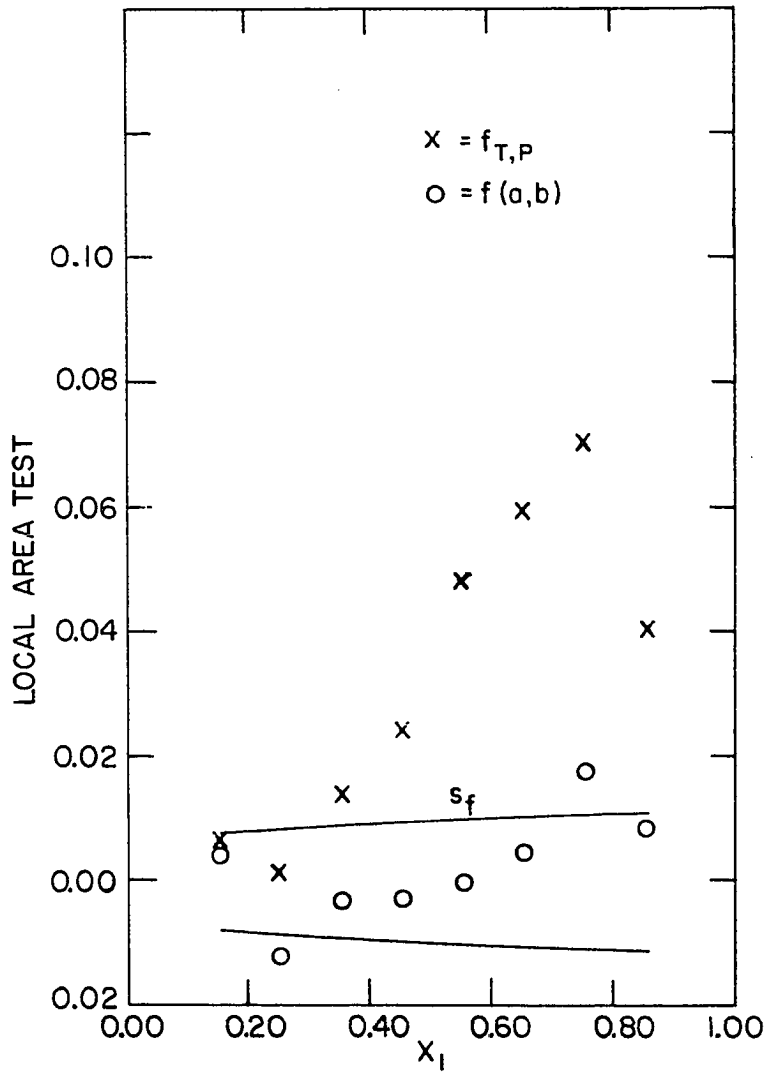

Figure 10. The effect of correcting $f_{T,P}$ for volume change of mixing

f(a,b) points are inside the confidence region. Repeating this analysis with other values of M ranging from 9 to 49 gives similar results. The precise description of $\Delta V/RT$ is not obtained from the regression analysis but the corrected values of f(a,b) usually satisfy the appropriate statistical criterion. The estimate of ΔH or ΔV seems to improve as the number of data points, M, increases. However, the distribution of the corrected f(a,b) values does not improve, apparently because of the narrower integration intervals obtained for larger values of M and the attendant smaller values of $T_b - T_a$ and $P_b - P_a$.

4. Discussion

The equations developed here are based on the approximate law of the propagation of error. The simulation of hypothetical systems and imposed random error indicates that s_f and s_γ are statistically correct. The equations developed by Tao (49) and by Chang and Lu (8) are based on a maximum error concept and are not readily interpreted in statistical terms. A comparison of the form of the s_f and s_γ equations with those of the above authors will illustrate the differences.

Equation 26 for $E(\gamma_i)/\gamma_i$ was developed by Tao as the equivalent of equation 75 for s_{γ_i}/γ_i. Consider a case where, e.g., $E(x_i) = s_{x_i}$, and

$$s_{x_i}/x_i = s_{y_i}/y_i = s_P/P = (d\ln P_i/dT)s_T = U.$$

Equation 75 would give, for this situation,

$$s_{\gamma_i}/\gamma_i = 2U$$

while equation 26 would give

$$E(\gamma_i)/\gamma_i = 4U.$$

Equation 26 is, therefore, not statistically correct.

Equation 32 for D was presented by Chang and Lu (8) as a means of estimating the error in d of equation 30. These expressions, D and d correspond to the terms s_f and f. Equation 32 was also developed in terms of maximum errors. The pressure terms, for example, in s_f and D are easily shown to differ by a factor of $\sqrt{2}$ and indicates that D is not easily interpreted in statistical terms. Further, Chang and Lu suggest that values of d which lie outside the error bound, $\pm D$, are inconsistent. Actually all values of d which are nonzero are inconsistent and a different terminology would appear to be more meaningful.

The terms consistent and inconsistent could well be reserved to apply only to representation equations. Representation equations which satisfy the Gibbs-Duhem equation identically are consistent. If they do not satisfy the Gibbs-Duhem equation, they are inconsistent. With this terminology, experimental data will always be inconsistent. A consistency test then determines only the accuracy of the data relative to the uncertainty of the measurements.

D. Characterization of the Activity Coefficient Error

The simulation of hypothetical systems as described above yields two sets of activity coefficients, γ_i and γ_i^e. The imposed random error in the measured variables induces an error in γ_i defined as

$$E(\gamma_i) = 100(\gamma_i^e - \gamma_i)/\gamma_i^e. \tag{85}$$

The per cent error in the activity coefficients, $E(\gamma_i)$, is related to

100 s_{γ_i}/γ_i in the same manner as f(a,b) is related to s_f. The hypothetical systems described in Tables 2 and 3 are used to verify and illustrate this relationship.

1. Verification of equations

The activity coefficient γ_i^e is calculated from products and ratios of terms containing normally distributed error. The existence of an approximately normal distribution in γ_i^e should therefore be verified. The procedure used is identical to that used for the consistency test except that each variable is evaluated at a point instead of over an interval. Again 200 iterations with M = 19 are used to calculate values of $\{\Sigma E(\gamma_i)^2/199\}^{\frac{1}{2}}$. This result should approximate 100 s_{γ_i}/γ_i as obtained from equation 75. Values of p_{γ_i}, defined as the per cent of the total number of $E(\gamma_i)$ values which lie inside the 68 per cent confidence region, are calculated and compared to the expected values. Table 6 shows the results obtained for system III.

Two values of p_{γ_i} exceed the fiducial limits of 68.3 ± 6.5 associated with the 1 in 20 probability level. Of the total runs made almost exactly one value in twenty was outside this range. This is the expected result for a variable containing only normally distributed random error.

The results in Table 6 are selected values obtained from runs with M = 19. Values of p_{γ_i} based on these 19 points were calculated for each of the 200 iterations and the average values and calculated standard deviations are presented in Table 7. Both the average values and the standard deviations compare very favorably with the expected

Table 6. Test for normal distribution in $E(\gamma_i)$

Imposed error	$\{\frac{\Sigma E(\gamma_1)^2}{199}\}^{\frac{1}{2}}$	$100\frac{s_{\gamma_1}}{\gamma_1}$	p_{γ_1}	$\{\frac{\Sigma E(\gamma_2)^2}{199}\}^{\frac{1}{2}}$	$100\frac{s_{\gamma_2}}{\gamma_2}$	p_{γ_2}
		$x_1 = 0.10$				
s_x=0.004,	4.25	4.00	66.0	0.47	0.44	65.0
s_y=s_P=s_T=0	3.83	4.00	69.5	0.43	0.44	68.5
s_y=0.004,	3.34	3.48	68.0	0.44	0.45	68.0
s_x=s_P=s_T=0	3.45	3.48	68.5	0.44	0.45	68.5
s_P=2.0,	0.40	0.44	77.0[a]	0.40	0.44	77.0[a]
s_x=s_y=s_T=0	0.45	0.44	71.0	0.45	0.44	71.0
s_T=0.2,	0.94	1.01	74.5	0.69	0.73	74.5
s_x=s_y=s_P=0	1.12	1.01	64.5	0.81	0.73	64.5
s_x=s_y=0.004,	5.25	5.43	69.0	1.07	1.07	70.0
s_P=2.0, s_T=0.2	5.07	5.43	73.5	1.07	1.07	68.0
		$x_1 = 0.55$				
s_x=0.004,	0.73	0.73	68.5	0.87	0.89	67.5
s_y=s_P=s_T=0	0.77	0.73	67.0	0.94	0.89	67.0
s_y=0.004,	1.61	1.58	66.5	0.54	0.54	66.5
s_x=s_P=s_T=0	1.69	1.58	65.5	0.56	0.54	65.5
s_P=2.0,	0.46	0.49	73.0	0.46	0.49	73.0
s_x=s_y=s_T=0	0.50	0.49	62.5	0.50	0.49	62.5
s_T=0.2,	1.12	1.01	63.5	0.82	0.73	63.5
s_x=s_y=s_P=0	1.03	1.01	63.5	0.75	0.73	63.5
s_x=s_y=0.004,	2.17	2.07	67.5	1.34	1.36	67.5
s_P=2.0, s_T=0.2	2.07	2.07	67.5	1.28	1.36	73.0

[a]The probability is less than 1 in 20 of obtaining numbers this large.

values of 68.3 and 10.7, respectively. Similarly, the results for M = 19 and M = 39 are very close to the expected values. The distribution of

Table 7. Average values of p_{γ_i} and the standard deviation for p_{γ_i}

Imposed error	Av. p_{γ_1} per cent	$\{\frac{\Sigma(p_{\gamma_1}-68.3)^2}{199}\}^{\frac{1}{2}a}$	Av. p_{γ_2} per cent	$\{\frac{\Sigma(p_{\gamma_2}-68.3)^2}{199}\}^{\frac{1}{2}a}$
$s_x=0.004$,	68.6	10.8	68.8	10.8
$s_y=s_p=s_T=0$	66.2	11.8	66.1	11.4
$s_y=0.004$,	68.6	10.9	68.6	10.9
$s_x=s_p=s_T=0$	67.2	11.5	67.2	11.5
$s_p=2.0$,	67.6	10.4	67.7	10.4
$s_x=s_y=s_T=0$	67.8	11.2	67.8	11.2
$s_T=0.2$	68.3	10.5	68.3	10.5
$s_x=s_y=s_p=0$	68.3	9.9	68.2	10.0
$s_x=s_y=0.004$	68.4	10.4	69.0	10.3
$s_p=2.0$, $s_T=0.2$	69.2	11.1	67.5	10.2
Average	68.0	10.9	67.9	10.7

[a] From equation 76: $s_{p_\gamma} = \{68.3(31.7)/19\}^{\frac{1}{2}} = 10.7$.

errors in the activity coefficients is, therefore, correctly described by the confidence regions predicted for normally distributed errors.

The approximate propagation of error formula, equation 57, is considered to be accurate if the coefficient of variation, defined as the standard deviation divided by the mean, is less than 0.2 for each of the measured variables. This approximate criterion of accuracy is verified by the results obtained for $E(\gamma_i)$ and s_{γ_i}/γ_i at low values of x_1. When the values of s_x/x_1 or s_y/y_1 become greater than 0.2, incorrect results are obtained. Thus the lowest values of x_1 and y_1 which can be reliably investigated are those for which $s_x < 0.2x_1$ and $s_y < 0.2y_1$. This causes no problem in the present method, however,

because with M = 99 and $s_x = s_y = 0.004$, for example, only two of the 99 points violate these conditions.

2. Effect of random error on s_{γ_i}

The hypothetical systems presented in Tables 2 and 3 are used to illustrate the effect of random measurement errors on the uncertainty in the activity coefficient. Equation 75 is used to calculate the variance in the activity coefficient. The variances in the measured variables that were used for illustrating the consistency test are also used here to illustrate the behavior of equation 75.

The 68 per cent confidence regions for $E(\gamma_1)$ are shown in Figure 11 for the four hypothetical systems. Separate curves are shown for $s_x = 0.004$, $s_y = 0.004$, $s_P = 2.0$ and $s_T = 0.2$ to illustrate the effect of error in each variable independently. These curves show why it is extremely difficult to determine γ_1 accurately for low concentrations of component one. The effect of uncertainties in x_1 and y_1 are magnified by the factors $1/x_1$ and $1/y_1$ so that the uncertainty in γ_1 approaches infinity as x_1 and y_1 approach zero. Techniques which could measure x_1 with a constant percentage error so that s_{x_1}/x_1 is constant would be superior in this case because s_{γ_1}/γ_1 would also be constant as shown by equation 75. Figure 11 also shows that the temperature and pressure errors cause a relatively constant uncertainty in the activity coefficient.

Comparison of Figures 11 and 7 indicates that the consistency test does not necessarily reflect the uncertainty in the activity coefficient. Figure 7 shows that measurement errors in x_1 described

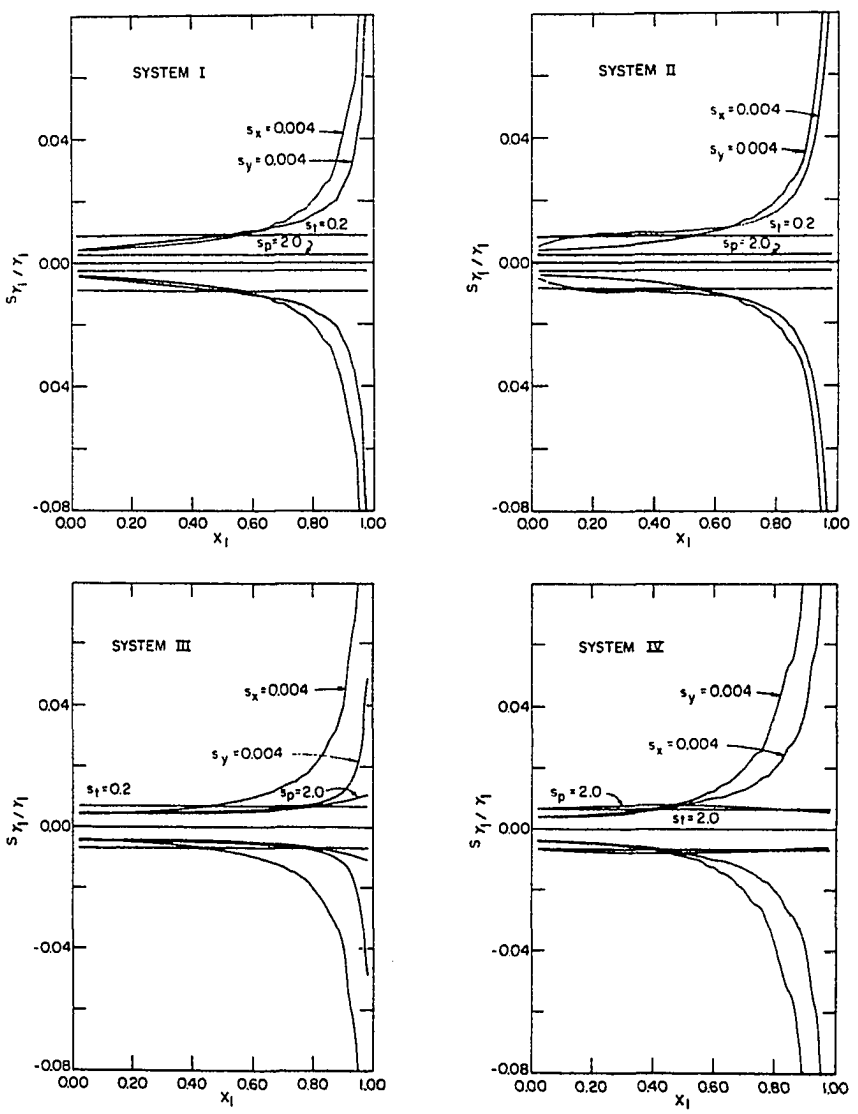

Figure 11. The effect of measurement uncertainties in only one variable on the confidence region for γ_1

by $s_x = 0.004$ will give very small values of f(a,b). If the uncertainty in the other measured variables is small, then the consistency test would indicate very accurate measurements. Figure 11 shows, however, that the uncertainty in the activity coefficient could still be very large. Conversely, uncertainty in measurement of temperature and pressure may cause large values for f(a,b) but contribute only small uncertainty to the activity coefficient. These comparisons illustrate the importance of using a consistency test only to check for systematic error. The value of a consistency test is greatly reduced if estimates of the uncertainty of the measurements are not known. In any event, the consistency test must be very carefully interpreted to obtain information about the accuracy of the activity coefficients.

The 68 per cent confidence regions for the activity coefficient of component two are shown in Figure 12 for the four hypothetical systems. The curves are quite similar to those for component one. The curves for $s_x = 0.004$ are identical for each system because x_1 is taken as the independent thermodynamic variable. The differences among the other curves for each system are due to the nature of the system. A general relationship for predicting the effect of uncertainty in y is as difficult in this case as it was for the consistency test equations. However, the effect at a particular value of x_1 is readily developed from equation 75 and an average effect could be calculated by evaluating each of the variables x_i, y_i, P and T at $x_1 = 0.5$, for example.

The effect of combined error in all four variables is obtained from Figures 11 and 12 by taking the positive square root of the sum

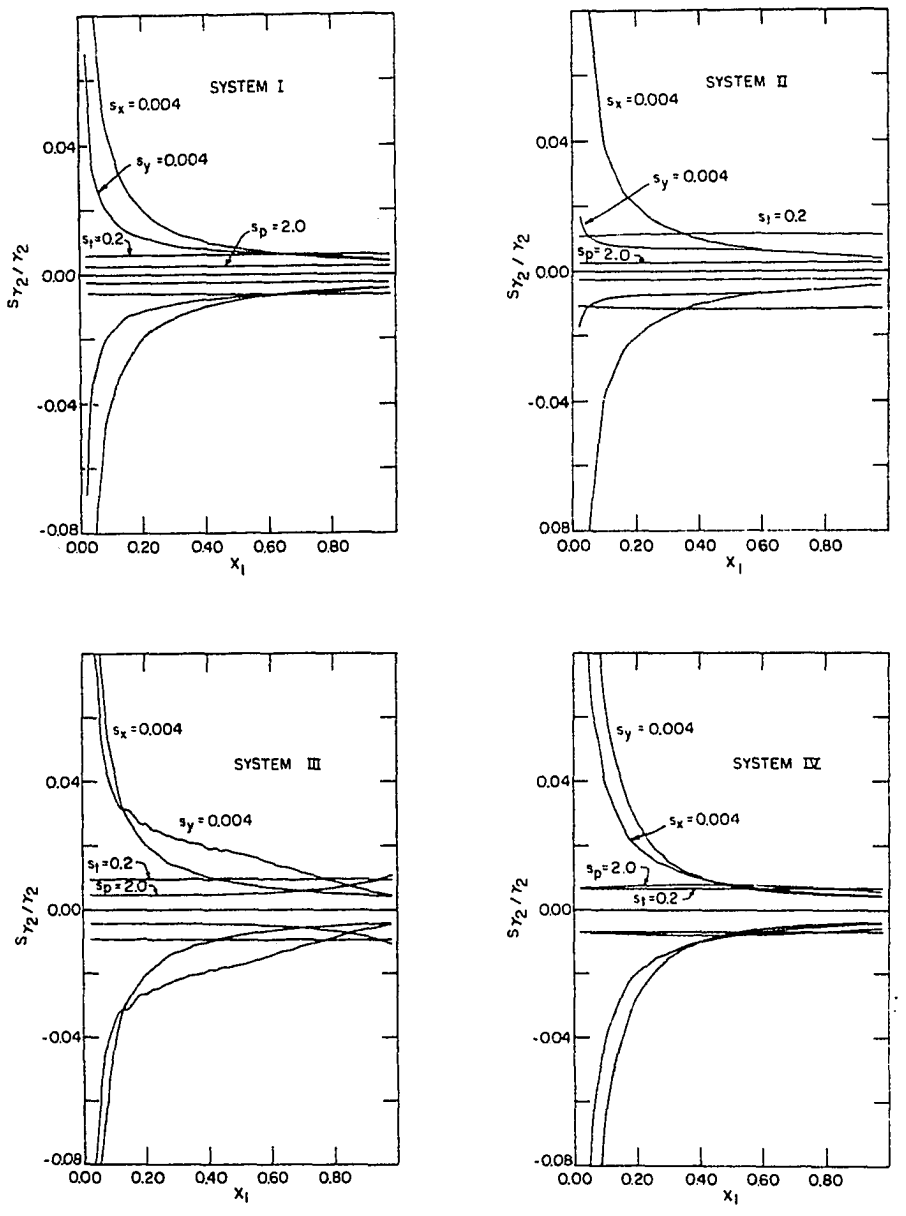

Figure 12. The effect of measurement uncertainties in only one variable on the confidence region for γ_2

of the squares of the individual contributions. These results are shown in Figure 13 for all four systems. Figure 13 shows only the standard deviations (in terms of the coefficient of variation, s_{γ_i}/γ_i) and not the confidence regions. The differences among these curves are quite small. The uncertainty in the measured variables causes about a 1 per cent uncertainty in the component one activity coefficients for each system at $x_1 = 0$. This uncertainty increases to between 2 per cent and 5 per cent at $x_1 = 0.8$. The uncertainty in the activity coefficient for component two is similar except the larger uncertainty occurs at small x_1.

The curves in Figure 13 are based on systems of fairly diverse nature for low pressure conditions as illustrated by the activity coefficient curves and x - y relations in Figures 4 and 5. The small differences in the standard deviations for the four systems indicates that the uncertainty in the activity coefficients is more dependent on the level of measurement error than on the nature of the system. Systems with very high volatility or extreme temperatures or pressures may, however, exhibit different characteristics.

E. Representation of Activity Coefficients

Activity coefficient data are usually most useful when expressed in terms of a representation equation. To accurately represent a set of activity coefficients, the form of the equation must be properly chosen and then the constants must be carefully determined. The proper form of an equation must be chosen to avoid misrepresentation caused by lack of fit. The evaluation of the constants then involves the

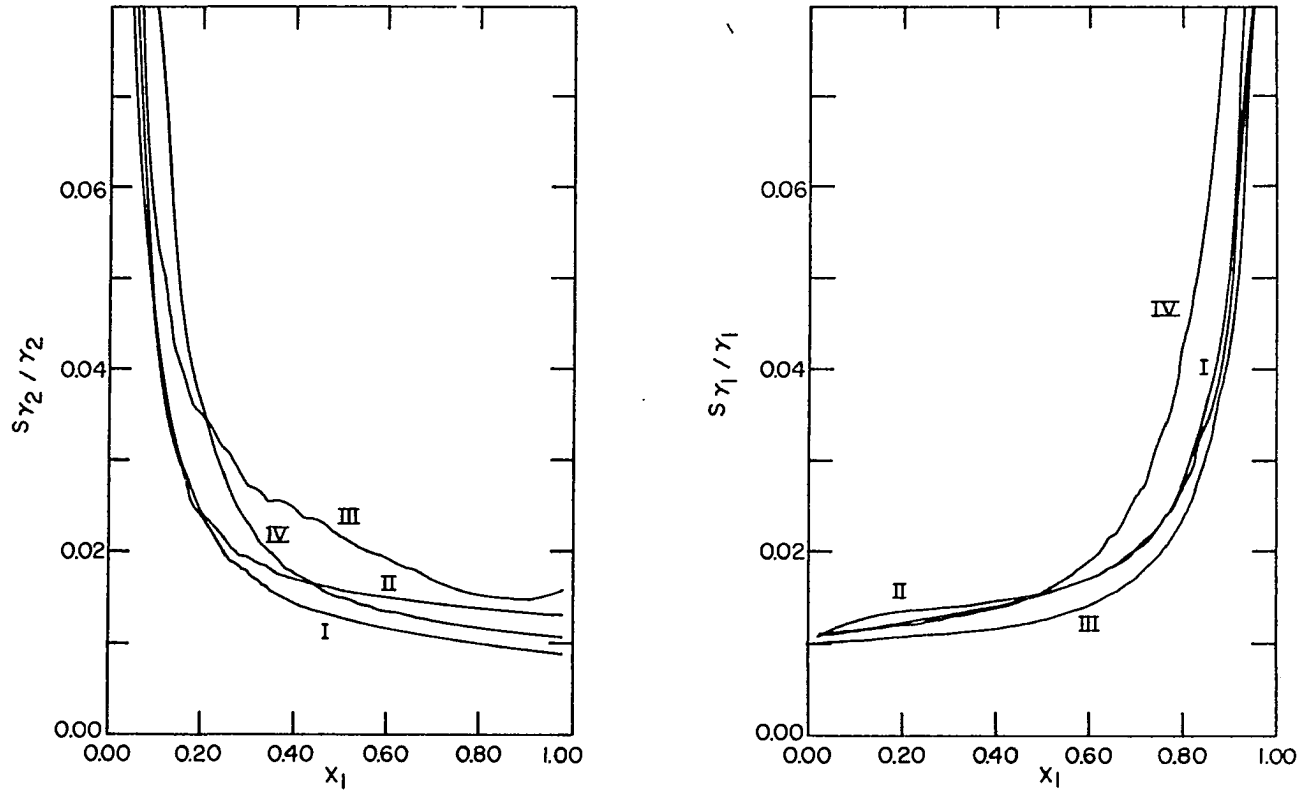

Figure 13. The effect of measurement uncertainty in all four variables on the uncertainty in the activity coefficients

smoothing of random errors. The best procedure to use in smoothing random error might be expected to depend on which measured variable contained the dominant error. The effect of measurement errors on the uncertainty associated with the determination of the constants in the representation equation is therefore investigated first. A method is then suggested for determining the lack of fit of the equation.

The activity coefficients are usually fit by least squares to an equation which guarantees consistent results. Three common functional forms for doing this are Q, dQ/dx_1 and Q/x_1x_2 as written in equations 44, 45 and 46. Recall that Q is defined as $Q = x_1 \ln\gamma_1 + x_2 \ln\gamma_2$ for a binary system. The Margules equation constants in these expressions are evaluated directly in a least-squares procedure by rewriting them as

$$Q = A(x_1 - 2x_1^2 + x_1^3) + B(x_1^2 - x_1^3), \tag{86}$$

$$dQ/dx_1 = A(1 - 4x_1 + 3x_1^2) + B(2x_1 - 3x_1^2), \tag{87}$$

and

$$Q/x_1x_2 = A(1 - x_1) + Bx_1. \tag{88}$$

The constant C in equations 44 to 46 is taken as zero to simplify the discussion.

1. Smoothing of random error

The hypothetical systems presented in Tables 2 and 3 are simulated by use of the Margules equations and random error is then imposed. A least-squares fit of the data by means of equations 86, 87 and 88 therefore involves no question of lack of fit and the smoothing of random error can be investigated.

The procedure followed is to select the system and the number of data points to be used. Then to simulate the system and impose random error on only one variable, e.g., x_1. A least-squares fit of the γ_1^e and γ_2^e data then yields values of A and B from each of the three functional forms, equations 86, 87 and 88. Another random error sample, described by the same variance, is chosen and values for A and B are again obtained. This procedure is repeated a sufficient number of times to estimate the standard deviations of A and B as obtained from each of the three equations.

The results from this procedure for systems I and II with M = 19 are shown in Table 8. Fifty iterations were used in all cases and the average values of A and B were usually correct to two decimal places. All three of the equations appear to smooth the data reasonably well. The largest standard deviation is 0.06 which indicates that 19 out of 20 times the constants will be determined to within ±0.12 and 2 out of 3 times they will be determined to within ±0.06. For a constant of unity this corresponds to 12 and 6 per cent errors, respectively. However, by choosing the appropriate functional form to fit the data (equations 86, 87 or 88) this uncertainty may be reduced by more than an order of magnitude. Note that the value of A in the Margules equations, 43, is the value of $\ln\gamma_1$ at $x_1 = 0$ and that $B = \ln\gamma_2$ at $x_1 = 1$. Thus the uncertainties in A and B are the uncertainties in the logarithms of the activity coefficients at $x_1 = 0$ and 1.

The choice among equations 86, 87 or 88 in fitting activity coefficient data depends on which variable contains the dominant error. With error only in x_1, the results in Table 8 show that Q and Q/x_1x_2 are

Table 8. Standard deviations of the Margules equation constants A and B

Imposed error	Q		dQ/dx_1		Q/x_1x_2	
	s_A	s_B	s_A	s_B	s_A	s_B
			System I			
s_x=0.004	0.000	0.000	0.027	0.029	0.001	0.001
s_y=0.004	0.009	0.006	0.019	0.023	0.018	0.013
s_P=2.0	0.009	0.010	0.000	0.000	0.014	0.014
s_T=0.2	0.028	0.024	0.002	0.002	0.048	0.041
Combined	0.025	0.022	0.033	0.040	0.051	0.047
			System II			
s_x=0.004	0.000	0.000	0.022	0.032	0.001	0.001
s_y=0.004	0.013	0.008	0.015	0.027	0.030	0.016
s_P=2.0	0.008	0.008	0.000	0.000	0.017	0.018
s_T=0.2	0.031	0.040	0.002	0.002	0.042	0.055
Combined	0.035	0.036	0.028	0.047	0.064	0.069

superior to dQ/dx_1. The differences are not as great when the only error is in y but the use of Q appears to give the best smoothing. Uncertainty in either temperature or pressure is always best smoothed by using dQ/dx_1. This is because $dQ/dx_1 = \ln\gamma_1/\gamma_2$ and errors in pressure and temperature tend to cancel. Therefore, if the dominant uncertainties are in temperature and pressure, dQ/dx_1 should be used to smooth the data. Q should be used to smooth uncertainties in x or y. When error occurs in all four variables, either Q or dQ/dx_1 may be used. In the

systems tested here the functional form Q/x_1x_2 always gave the poorest results in fitting activity coefficient data.

The standard deviations, s_A and s_B, are essentially linear in s_x, s_y, s_P and s_T, individually. That is, if s_y is doubled, s_A and s_B are approximately doubled while s_A and s_B are zero if s_y is zero. The values of s_A and s_B do not appear to depend strongly on the system studied. The conclusions reached here should, therefore, be useful in estimating the uncertainty in Margules constants obtained from a least-squares fit. Other types of equations such as the Van Laar and Wilson equations were not considered. However, the results of a least-squares fit would seem to depend more on the functional form used than on the particular type of equation. The forms Q and dQ/dx_1 would therefore appear to be the most desirable forms to use.

2. Selection of equations

A least-squares fit of activity coefficient data via equations 86 or 87 provides two constants, A and B. These constants may be used in equations 43a and 43b to calculate activity coefficients which satisfy the Gibbs-Duhem equation identically. Equations 43a and 43b are therefore a consistent pair of equations. However, they may represent a set of experimental data quite poorly. The accuracy of the representation can be checked by calculating an error term similar to that in equation 85, viz.,

$$E(\gamma_i^r) = 100(\gamma_i^r - \gamma_i)/\gamma_i^r. \tag{89}$$

The superscript r indicates an activity coefficient obtained from equations 43 by using the value of x_i associated with the data point γ_i.

Equation 89 then gives the per cent error in the activity coefficient at every point for which a value of γ_i is available. If the Margules equations correctly describe the liquid phase behavior as reflected in the activity coefficients, only random errors will be present. If the Margules equations are incorrect, nonrandom trends will appear in plots of $E(\gamma_i^r)$ versus x_i.

The uncertainty of the measurements associated with a set of vapor-liquid equilibrium data is seldom known. However, the consistency test can be used to estimate these uncertainties. The uncertainty of the measurements can then be used in equation 75 to provide a confidence region for $E(\gamma_i^r)$. The ethanol-chloroform data reported by Scatchard and Raymond (44) are used to illustrate the procedure.

Scatchard and Raymond reported 25 data points for the ethanol-chloroform system at 45^0C. At each point the values of x_1, y_1 and P were reported. The vapor pressures of the pure components were also reported and equation 1, with $\theta_i = 1$, was used to calculate the activity coefficients. Equation 48 with N = 2, $T_b = T_a$ and $\Delta V = 0$ was used to calculate f(a,b) for each of the 24 intervals. The results are plotted against the mole fraction of ethanol (x_1) in Figure 14.

The first step in a consistency test procedure should be a check on the randomness of the distribution of f(a,b) about zero. In the present case,7 of the 24 values are greater than zero. Equation 78 with $N_+ = 7$ and M = 24 yields $\chi^2 = 4.17$. The probability is, therefore, less 0.05 that this distribution is the result of random measurement errors only. The trend in f(a,b) appears to be toward more negative

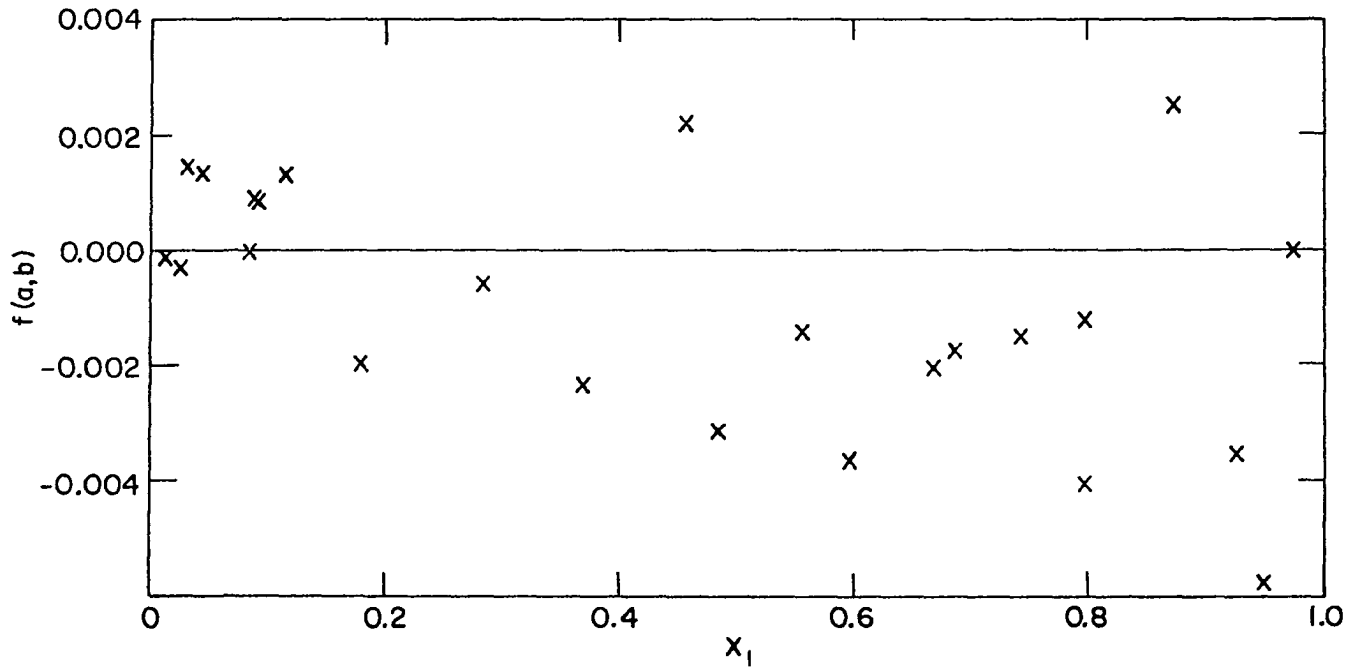

Figure 14. The consistency test results for the ethanol-chloroform system

values with increasing x_1. Mixon, Gumowski and Carpenter (34) have tested this data and concluded that the activity coefficients should be corrected for vapor phase nonideality. Their results were obtained from a complex polynomial fit and successive approximation procedure on a computer. The same conclusion is reached by the procedure here with much less effort.

The correction of the activity coefficients is not required for the present purposes and only random errors are assumed to be present. A confidence region is then established for which the data can be considered accurately measured. Obviously the neglect of errors which contribute a systematic trend to f(a,b) can not always be ignored. The present procedure indicates that an error is present in the data but it does not indicate the magnitude of the effect on the activity coefficients. The assumption that only random error is present may in some cases lead to erroneous conclusions. The assumption happens to be a reasonable one in this instance, however.

There is no unique combination of values for s_x, s_y, s_p and s_T which can be assigned purely on the basis of statistical considerations. However, in most cases, some knowledge about the experimental technique will assist in assigning reasonable values. The activity coefficients for ethanol and chloroform appear to plot quite smoothly so that the value of s_x is probably small. Reasonable estimates appear to be $s_x = 0.001$, $s_y = 0.001$, $s_P = 0.02$ and $s_T = 0.03$. The 50 per cent and 68 per cent confidence regions associated with these estimates are superimposed on the f(a,b) results in Figure 15. The values of p_f determined from Figure 15 are 75 per cent for the 68 per cent

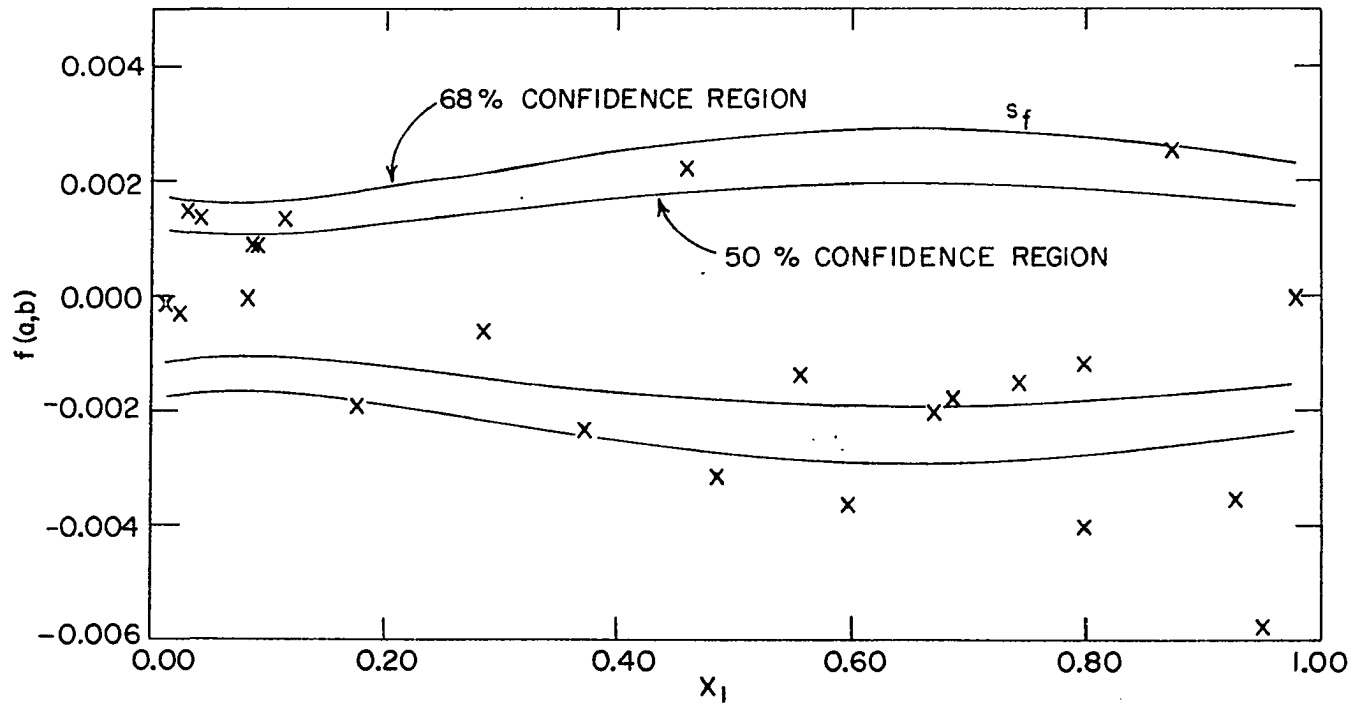

Figure 15. The consistency test results and the confidence regions for the ethanol-chloroform system

confidence region and 50 per cent for the 50 per cent confidence region. The data can therefore be considered accurate within the estimated uncertainties in the measured variables. Thus, even neglecting the effect of θ_i, the data are quite accurately measured. Other estimates for s_x, s_y, s_P and s_T could be used but the confidence region is reasonably well fixed by the data and there appears to be no need to estimate values of s_y, s_P and s_T which could individually contribute all the observed uncertainty in f(a,b).

The uncertainties s_x, s_y, s_P and s_T are then used in equation 75 to calculate s_{γ_i}/γ_i. The s_{γ_i}/γ_i values provide a confidence region for $E(\gamma_i^r)$. Equations 86 and 87 are used to fit the activity coefficient data. The two equations give nearly the same values for A and B. Therefore, only the results from equation 87 for dQ/dx_1 are discussed. Equations 43 with the constants A and B give values of γ_i^r and equation 89 is used to calculate $E(\gamma_i^r)$. The results for $E(\gamma_1^r)$ and the 68 per cent confidence region are plotted in Figure 16. The results for component two (chloroform) are shown in Figure 17.

The Margules equations do not fit the data with the accuracy justified by the consistency test. The most prominent error of 5 per cent in the component one results occurs near $x_1 = 0.5$. Although this error may be acceptable for some purposes, it would appear to negate the effort expended on the consistency test because the accuracy of representation is much poorer than the accuracy of measurement. The three constant Margules equations do not significantly reduce the lack of fit for this system. Some other representation equation or an interpolation procedure should be sought if the data are to be represented

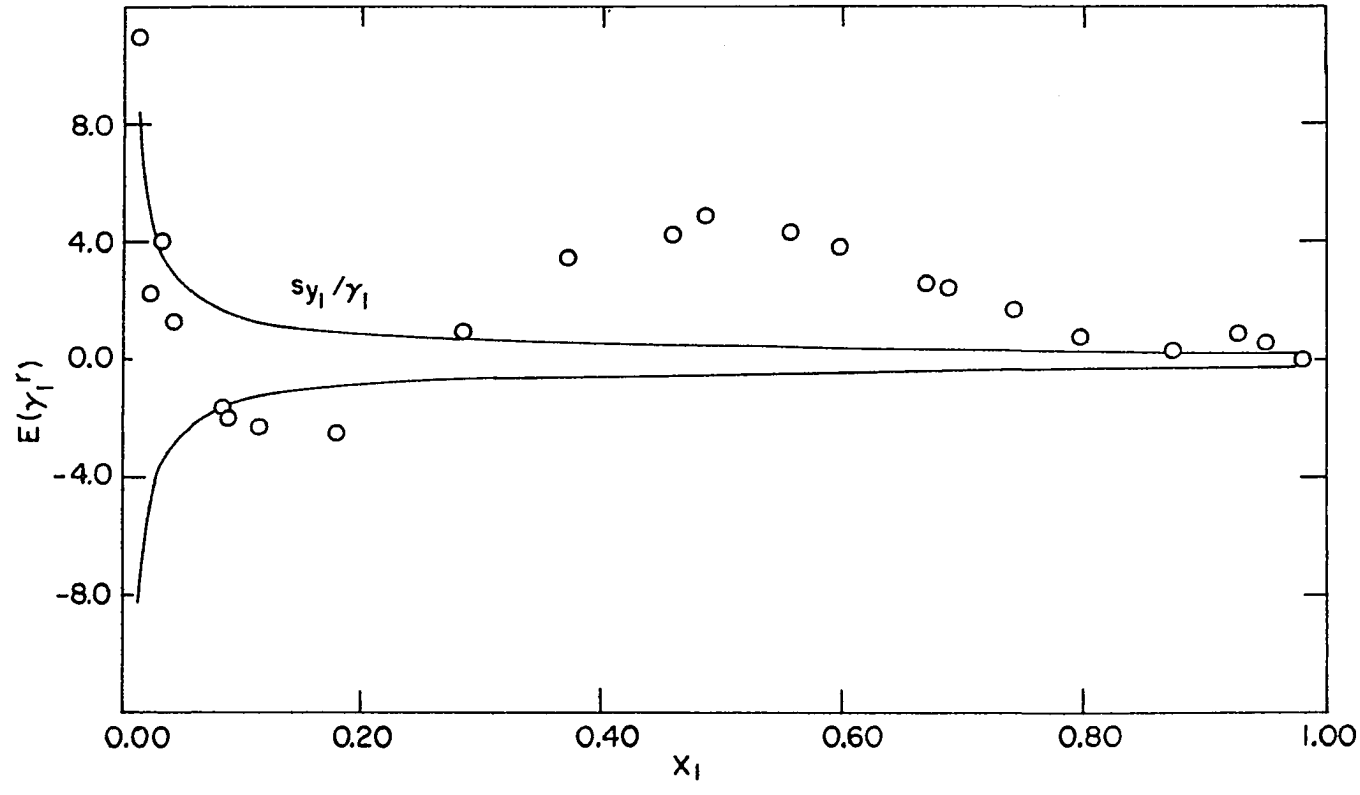

Figure 16. Analysis of the accuracy of a least-squares fit of the Margules equations used to represent the activity coefficient of ethanol

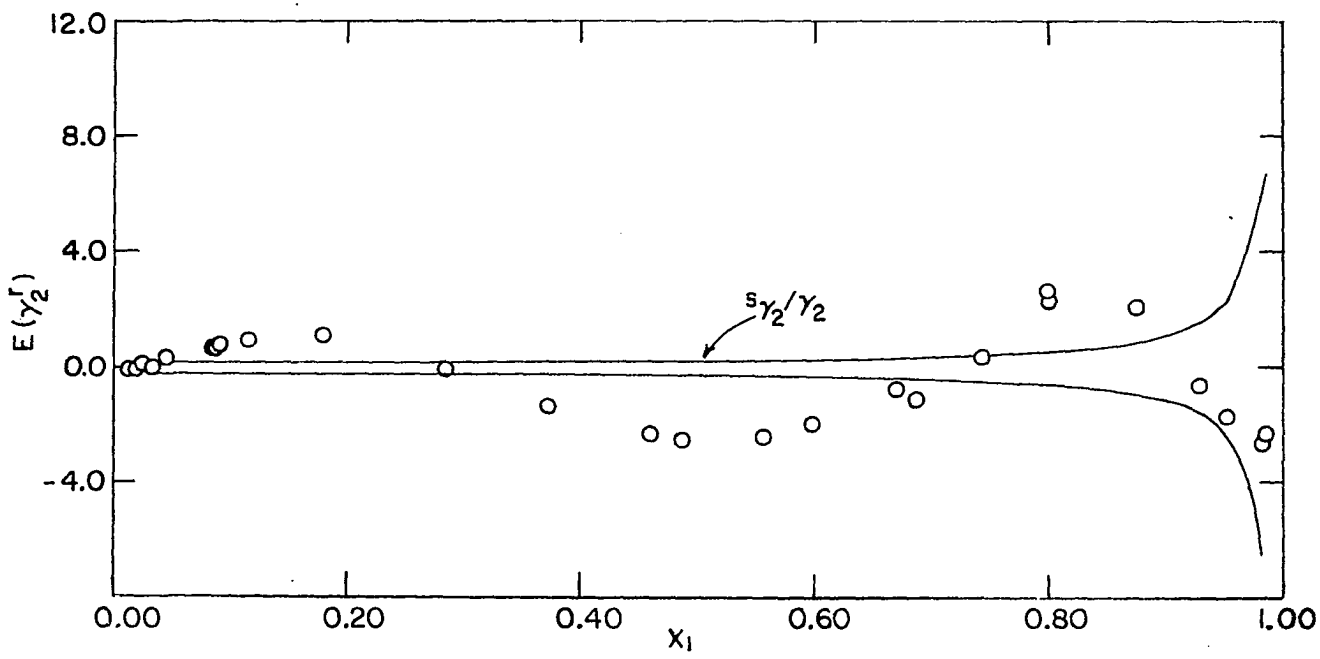

Figure 17. Analysis of the accuracy of a least-squares fit of the Margules equations used to represent the activity coefficient of chloroform

more accurately.

The ethanol-chloroform activity coefficient data, the Margules equation representation and a smooth curve drawn through the data are shown in Figure 18. Values of $E(\gamma_i^r)$ were also calculated from the smoothed curves in Figure 18 as a check on the accuracy of the confidence region. The results are shown in Figure 19. These results may reflect the inaccuracy of the graphical technique more than the uncertainty in the activity coefficient (particularly as γ_i approaches unity) but the distribution about the confidence region appears reasonable and tends to support the estimates of the measurement uncertainties. The values of p_{γ_1} and p_{γ_2} are 68 per cent and 48 per cent, respectively, for the 50 per cent confidence region.

The procedure illustrated in this section provides a useful purpose for the consistency test. A set of vapor-liquid equilibrium data can be described as accurate within given estimates of the experimental uncertainties and the accuracy of a representation equation can be checked. With careful graphical techniques a smooth curve through the activity coefficient data may also be helpful in estimating the experimental uncertainties in the activity coefficient.

F. Applications

The application of the procedures presented here is illustrated by considering two ternary systems and the six binary systems associated with them. The systems are the water-methanol-acetone system studied at a constant temperature of 100^0C by Griswold and Wong (14) and the n-octane-ethylbenzene-cellosolve system studied at constant pressure

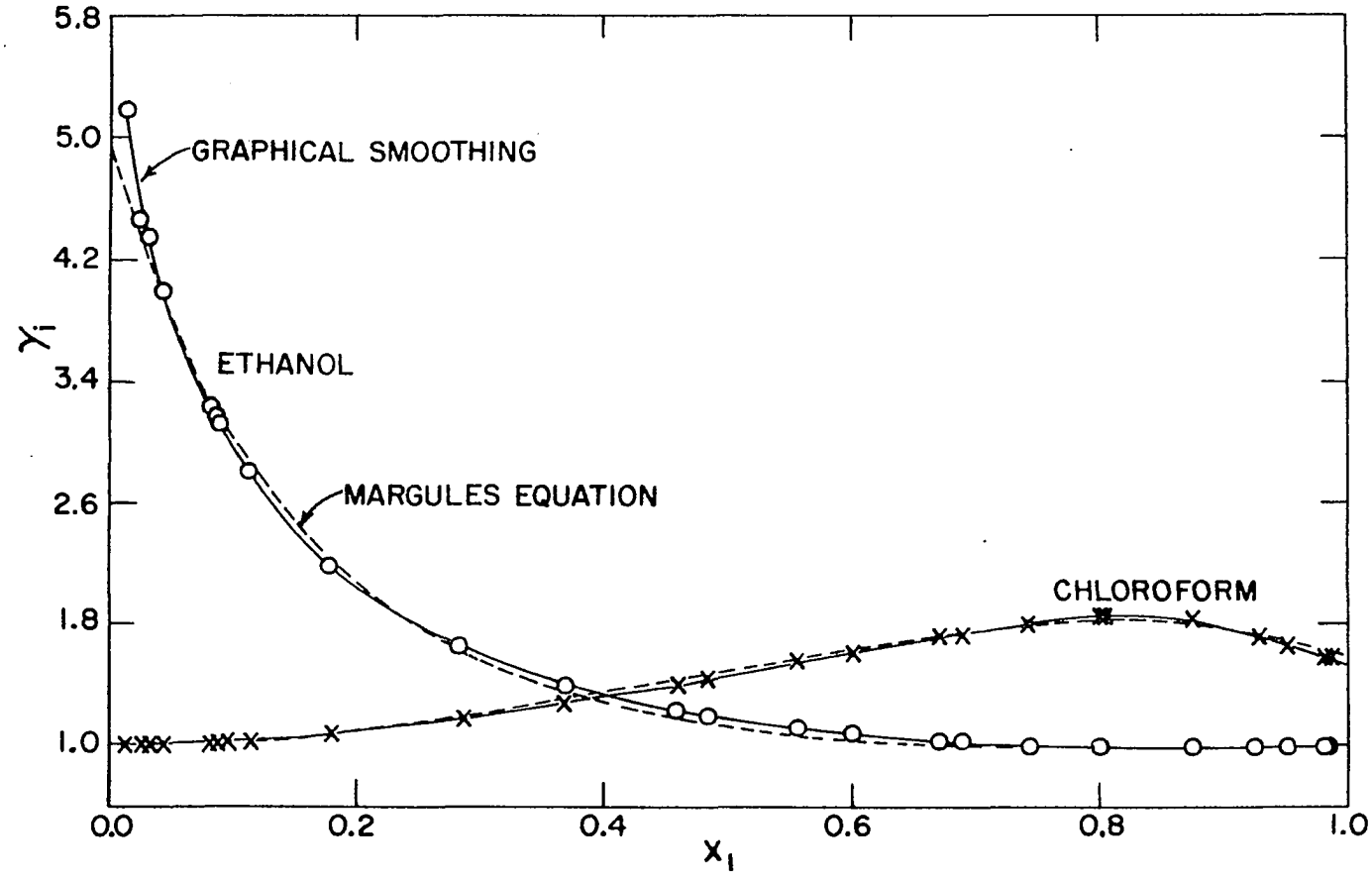

Figure 18. The activity coefficient data and representation for ethanol-chloroform (44)

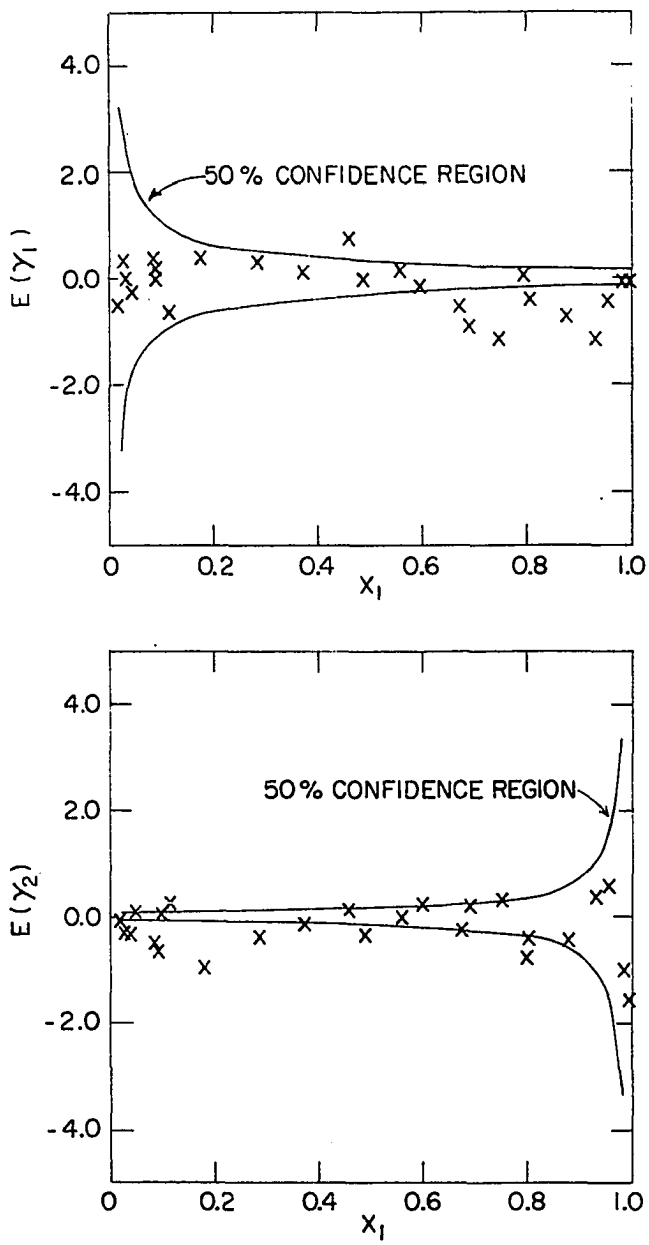

Figure 19. The error in the activity coefficients of ethanol-chloroform as indicated by graphical smoothing

of 760 mmHq by Murti and Van Winkle (35). Griswold and Wong presented binary data on acetone-water, methanol-water and acetone-methanol in the same article. The binary data for n-octane-cellosolve and ethylbenzene-cellosolve were reported by Murti and Van Winkle (35) and the n-octane-ethylbenzene data were reported by Yang and Van Winkle (61).

1. Water-methanol-acetone system (14)

The binary data associated with this ternary were reported as activity coefficients corrected for vapor phase nonideality. The Margules equations were fit to the three sets of binary data and a ternary interaction constant was obtained from the ternary data. The binary constants and the ternary constant which were reported are given in Table 9. Griswold and Wong indicated that the data were not well represented but the general trends were reproduced. The constants used are defined by equations 90 and 91. The binary equation is

$$\ln\gamma_1 = x_2^2\{A_{12} + 2x_1(A_{21} - A_{12})\}, \tag{90}$$

and the ternary equation is

$$\begin{aligned}\ln\gamma_1 = {} & x_2^2\{A_{12} + 2x_1(A_{21} - A_{12})\} + \\ & x_3^2\{A_{13} + 2x_1(A_{31} - A_{13})\} + \\ & x_2x_3\{A_{21} + A_{13} - A_{32} + 2x_1(A_{31} - A_{13}) + \\ & 2x_3(A_{32} - A_{23}) - C(1 - 2x_1)\}.\end{aligned} \tag{91}$$

The subscripts 1, 2, 3 refer to components of the mixture (1 = water, 2 = methanol, 3 = acetone). The equations for the other two components are obtained by the rotation principle, e.g., the equation for component two is obtained by replacing 1 by 2, 2 by 3 and 3 by 1 in each term.

Table 9. Binary constants for the Margules Equations[a]

Source	A_{12}	A_{21}	A_{13}	A_{31}	A_{23}	A_{32}
Griswold (14)	0.3569	1.0177	1.3309	2.2243	0.7046	0.3339
dQ/dx_1	0.3621	0.8441	1.2169	2.1681	0.7107	0.3874
Q	0.3533	0.8465	1.2055	2.1596	0.4451	0.3847
Q/x_1x_2	-0.0392	1.8445	0.7711	3.1270	0.4578	0.0894

[a]Component 1 = water, component 2 = methanol, component 3 = acetone.

The differences among the three functional forms, equations 86, 87 and 88, which can be used in a regression analysis is illustrated here by refitting the data by each of the three methods. The resulting constants are also shown in Table 9.

Griswold and Wong extrapolated a smooth curve through the activity coefficient data to x_i = 0.0 and 1.0 to obtain the constants they reported. Such a procedure necessarily weights the activity coefficients near the endpoints rather heavily. Table 9 shows a wide range of values for the constants. The activity coefficients calculated by using the pairs of constants obtained by Griswold and those obtained from least-squares fits in this work are compared to the data to illustrate some of the problems in choosing and fitting representation equations.

Figure 20 shows the acetone-methanol activity coefficient data plotted against the mole fraction of acetone. A smooth curve has also

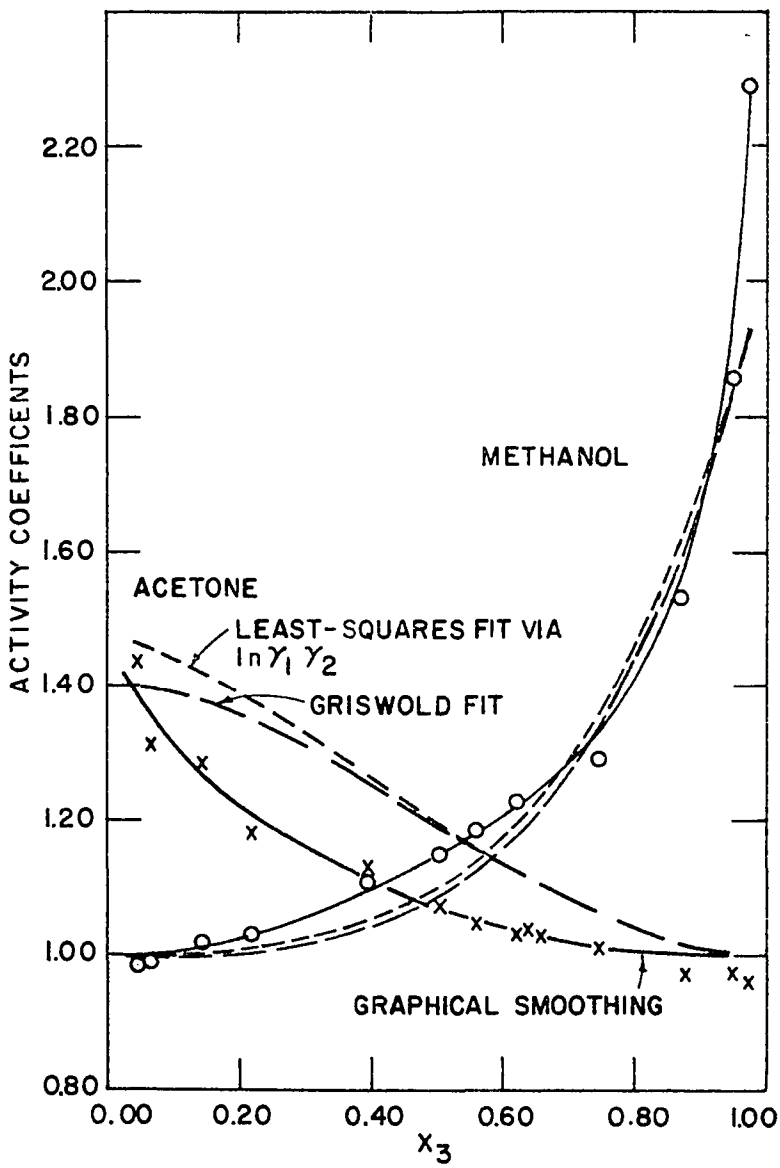

Figure 20. The acetone-methanol activity coefficient data and representations

been drawn through the data. The last three activity coefficient data points for acetone appear to be quite low. The activity coefficient of methanol does not show a maximum so the acetone activity coefficients should not be less than unity. The curves obtained by using equation 90, with the constants reported by Griswold and the constants obtained from a least squares fit of $\ln\gamma_1/\gamma_2$ are also shown. The Margules equations fit the data quite poorly in both cases. Figure 21 shows the same results for the methanol-water system. The representation equations are reasonably good in this case. The least-squares fit of dQ/dx_1 appears to be slightly better than the result obtained by Griswold. Figure 22 shows the results for the acetone-water binary. Again the least-squares fit is slightly superior and the representation is fair. All three systems show considerable scatter in the activity coefficient data as plotted against the liquid mole fraction.

The reason for the differences among the constants obtained by the different least-squares procedures is illustrated by plotting the data in the form used in the least-squares fit. The acetone-methanol data are used in this illustration. Figure 23 shows $\ln\gamma_3/\gamma_2$ plotted against x_3. The solid curve represents the result obtained from the Margules equations and the constants reported by Griswold and Wong. The dashed curve is the result from the least-squares fit of $dQ/dx_3 = \ln\gamma_3/\gamma_2$. Note that any values of A_{23} and A_{32} will automatically satisfy the area test. The curves in Figure 23 therefore enclose equal areas above and below 0.0 when extended to $x_3 = 0.0$ and 1.0. Obviously the Margules equations do not represent the data very well.

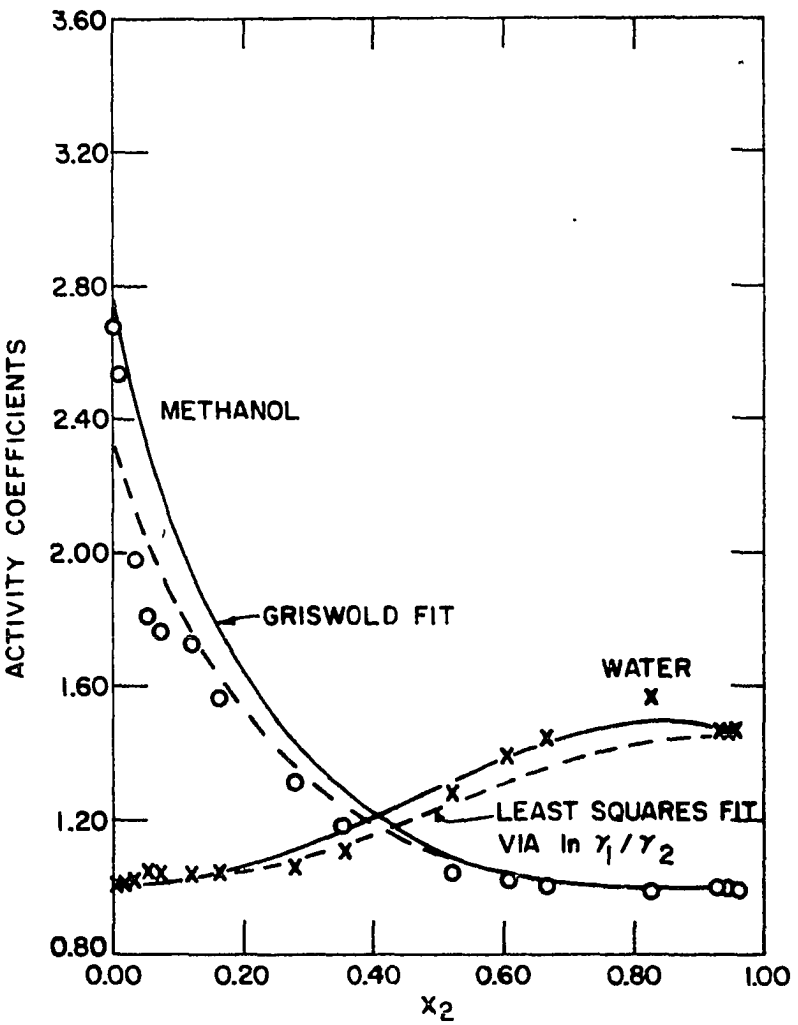

Figure 21. The methanol-water activity coefficient data and representations

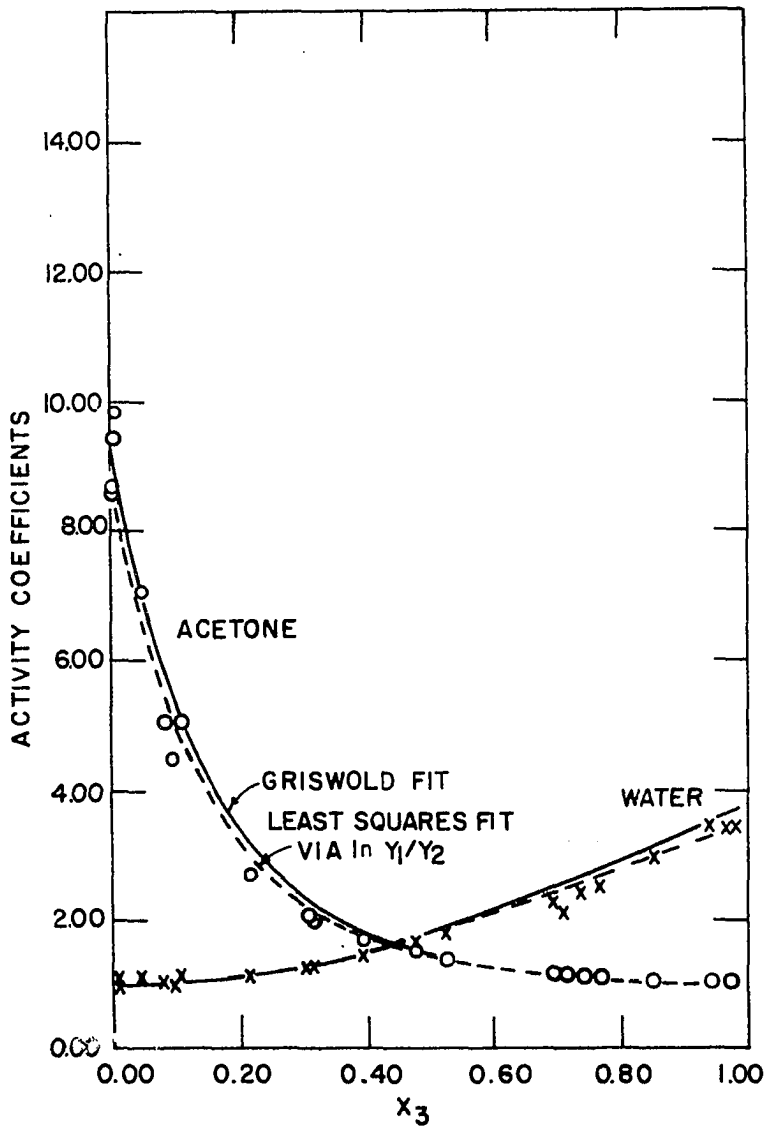

Figure 22. The acetone-water activity coefficient data and representations

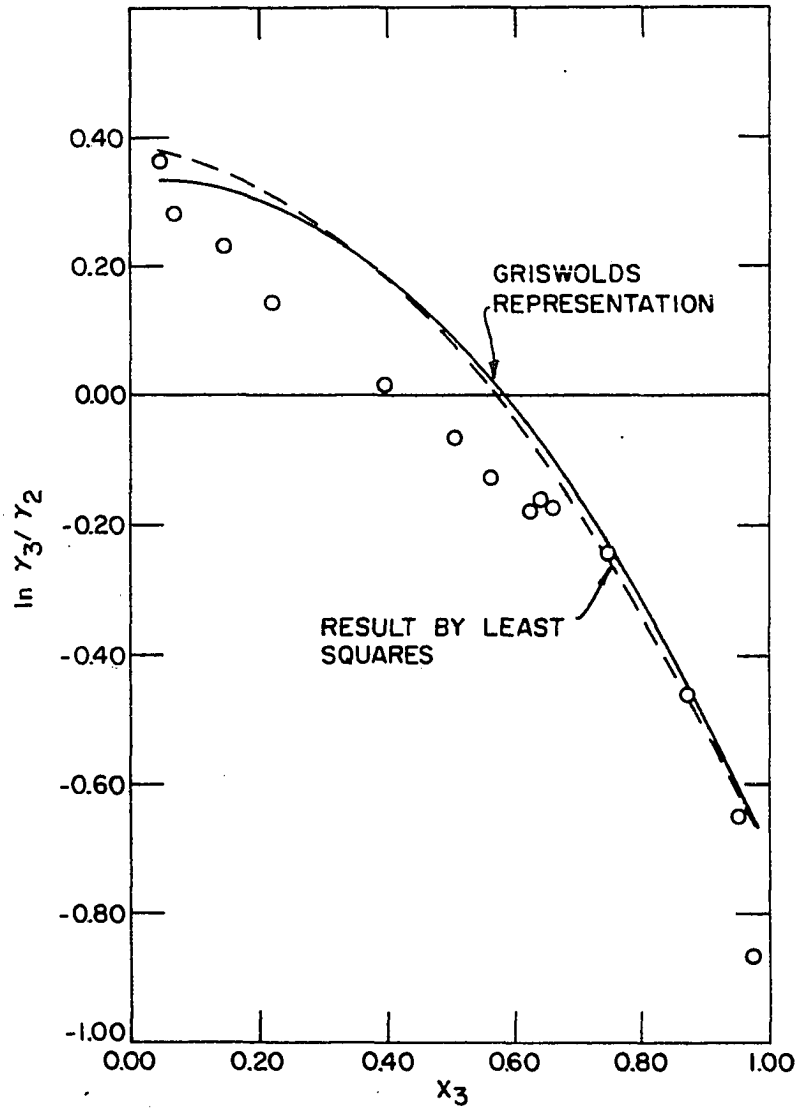

Figure 23. The acetone-methanol activity coefficient data represented by means of $\ln\gamma_3/\gamma_2$

Figure 24 shows similar results in the form of Q versus x_3. Here there is a substantial difference between the value of Q obtained by Griswold and that obtained by a least-squares fit. The activity coefficient curves as described by the constants obtained from fitting Q are shown in Figure 25. The fit of the acetone data is much better than that obtained from the $\ln\gamma_3/\gamma_2$ fit (shown in Figure 20). However, the methanol data are not as well represented by the constants obtained from Q as by those from $\ln\gamma_3/\gamma_2$.

Figures 26 and 27 show the results obtained from a least-squares fit of Q/x_3x_2. Figure 26 shows that the least-squares fit was heavily influenced by the data point at $Q/x_3x_2 = -0.74$. The curve resulting from Griswold's constants is little better, however. The activity coefficient curves are poorly represented by the least-squares fit as shown in Figure 27. The form Q/x_3x_2 appears to be too sensitive to reliably smooth the large uncertainties associated with these data. In fact, it is not obvious how a straight line should be drawn in Figure 26 to best represent the data. The Composition-Resolution test applied to Figure 26 would, therefore, yield activity coefficient curves quite different than the data and illustrates the danger of using the Composition-Resolution test to calculate a set of consistent data.

Figure 28 shows the results obtained by obtaining a least-squares fit of the Margules equations in each of the forms $\ln\gamma_2/\gamma_1$, Q and Q/x_2x_1 for the methanol-water system. The Margules equations represent the data somewhat better for the methanol-water system than they did for the acetone-methanol system. However, the Q/x_2x_1 fit is again

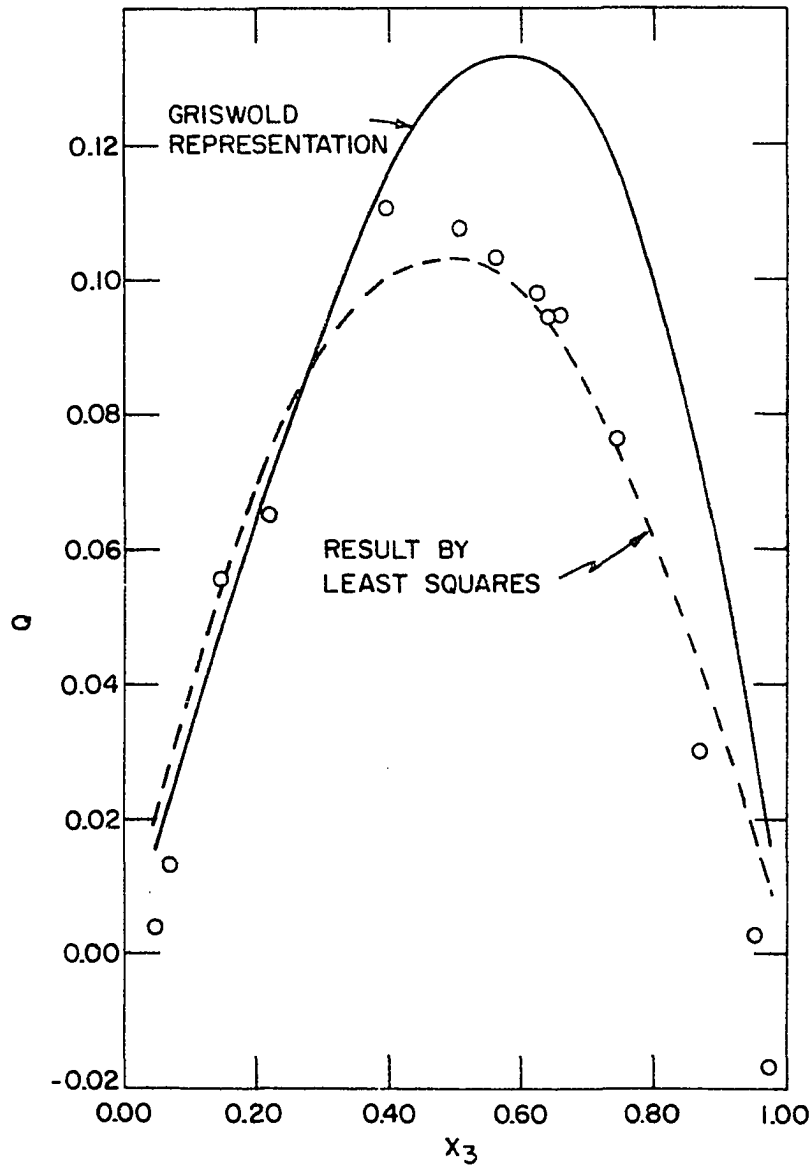

Figure 24. The acetone-methanol activity coefficient data represented by means of Q

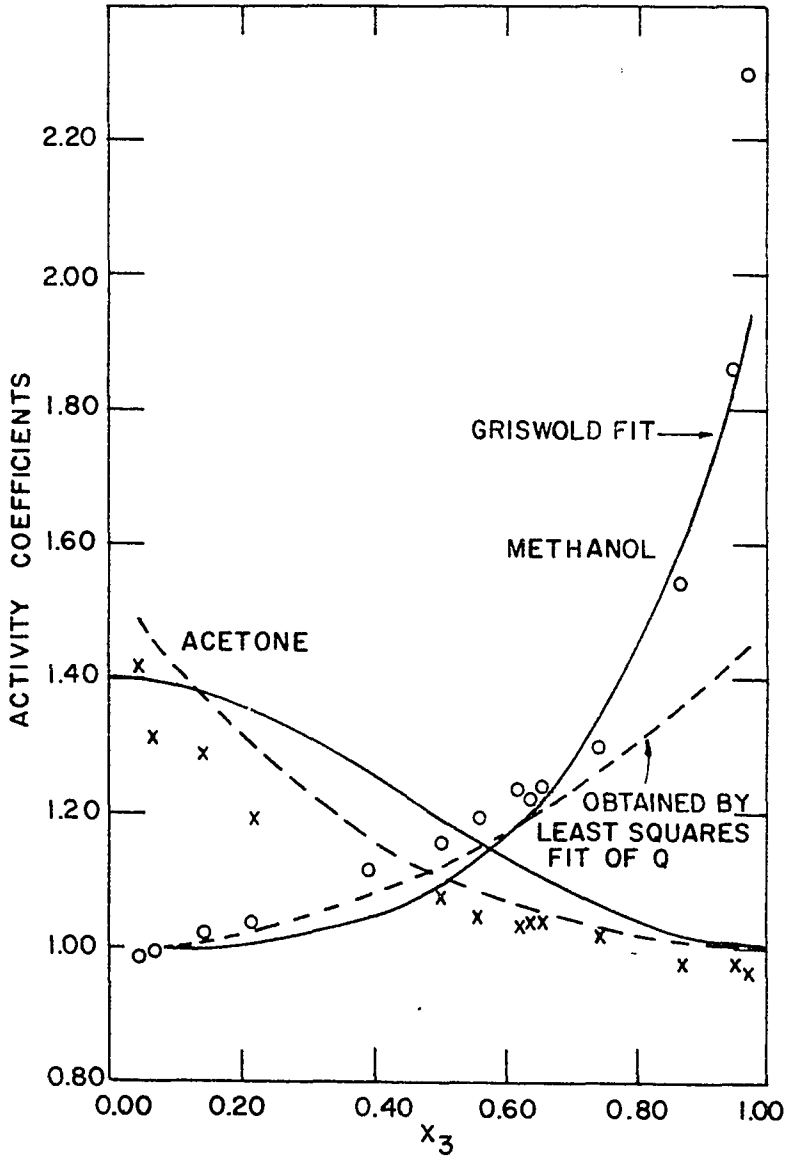

Figure 25. The acetone-methanol activity coefficient data and the representation obtained by means of Q

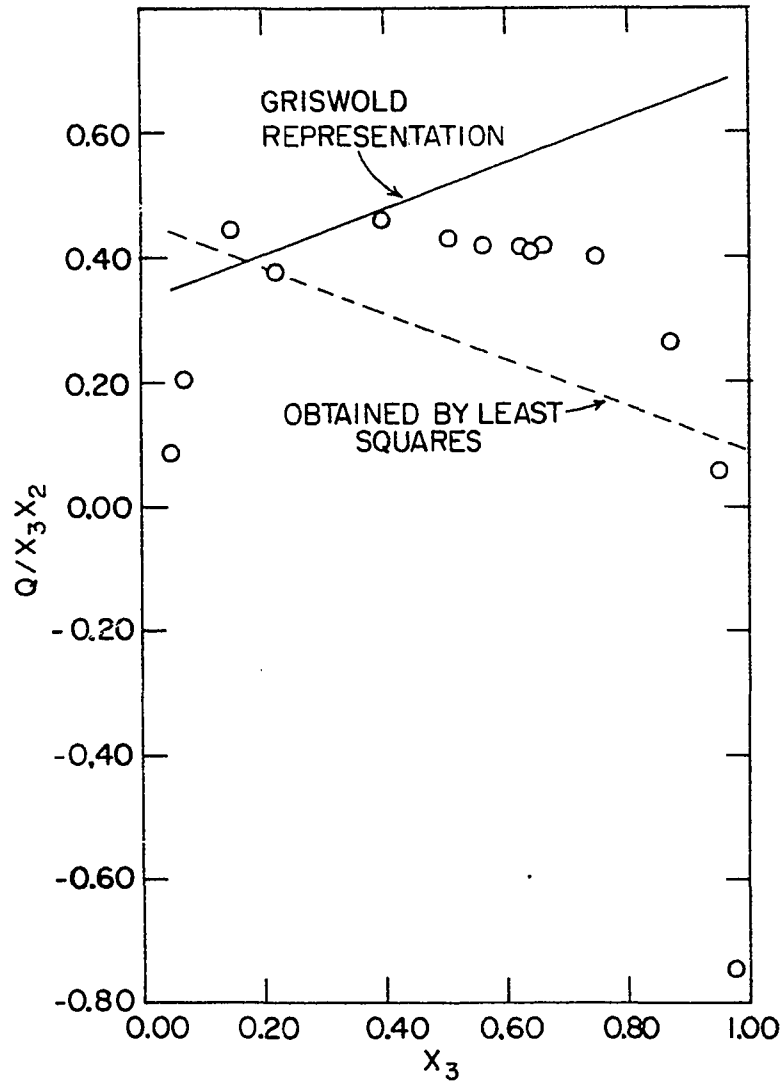

Figure 26. The acetone-methanol activity coefficient data represented by means of Q/x_3x_2

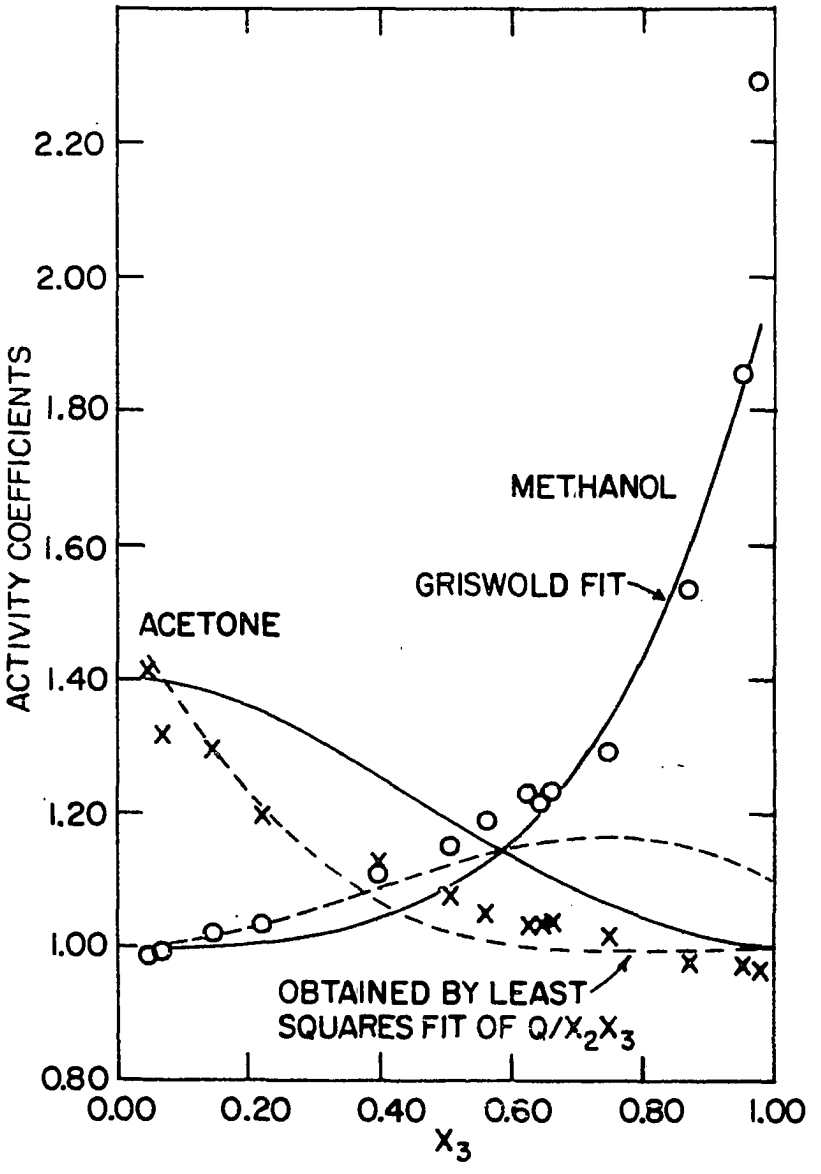

Figure 27. The acetone-methanol activity coefficient data and the representation obtained by means of Q/x_3x_2

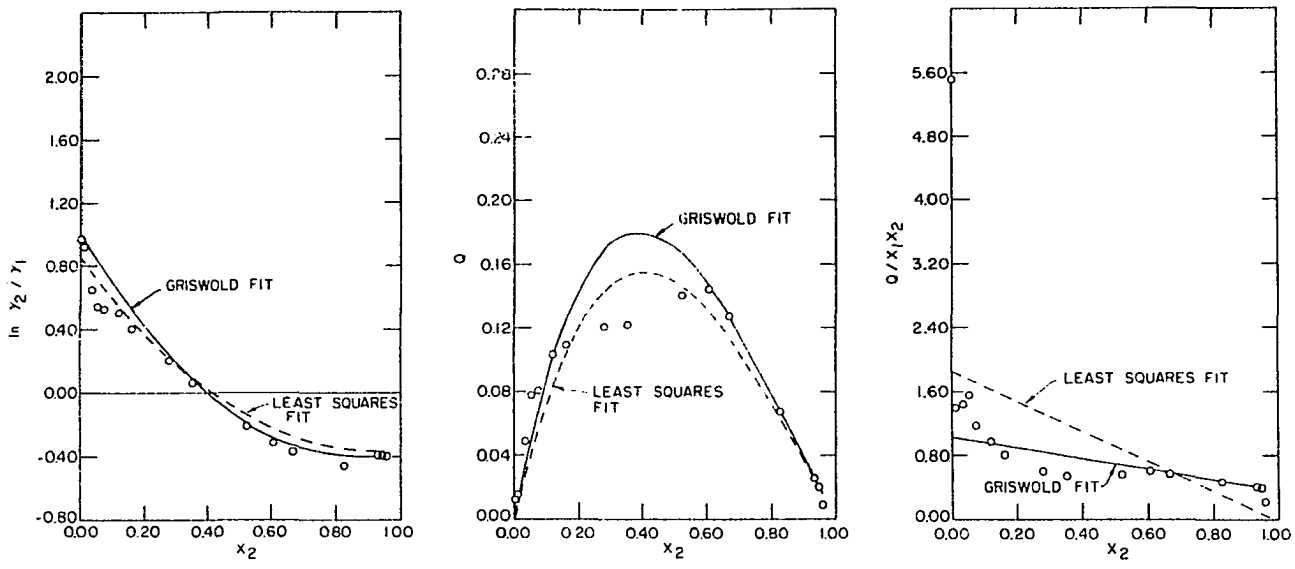

Figure 28. The results of least-squares fits for the methanol-water system

heavily influenced by one point and gives very poor results.

The consistency test was applied to the three sets of binary data and values of f(a,b) were calculated. Trapezoid rule corrections are not required because the integration intervals are usually less than 0.1 and the uncertainties of measurement are obviously quite large. The corrections are therefore negligible. Griswold and Wong reported their pressure measurements to be accurate to ±0.5 per cent. The pressure was approximately 50 psia so that s_p is taken as 0.3 psia. The analytical accuracy of composition measurement was reported as 2 per cent. This is interpreted to mean 2 per cent at $x_i = 0.5$ so that $s_x = 0.004$ and $s_y = 0.004$ is used. The uncertainty in temperature measurement was not reported. Two trial computer runs were sufficient to pick $s_T = 0.3^0C$ as a reasonable estimate of the uncertainty in the temperature measurement.

The consistency test results and the 68 per cent confidence region for f(a,b) are shown in Figures 29, 30 and 31 for the acetone-methanol system, the methanol-water system and the acetone-water system, respectively. The liquid compositions used in these plots are for the first named component. Fourteen data points were reported for the acetone-methanol systems and the value of p_f is 77 per cent. The confidence region in Figure 29 could be narrowed substantially without affecting the value of p_f. However, the measurement techniques were the same for each system so that the same uncertainties are used in defining the confidence region for each system. Sixteen data points were obtained for the methanol-water system and p_f is 80 per cent. Twenty-two data points were reported for the acetone-water system and p_f is 66.7 per cent.

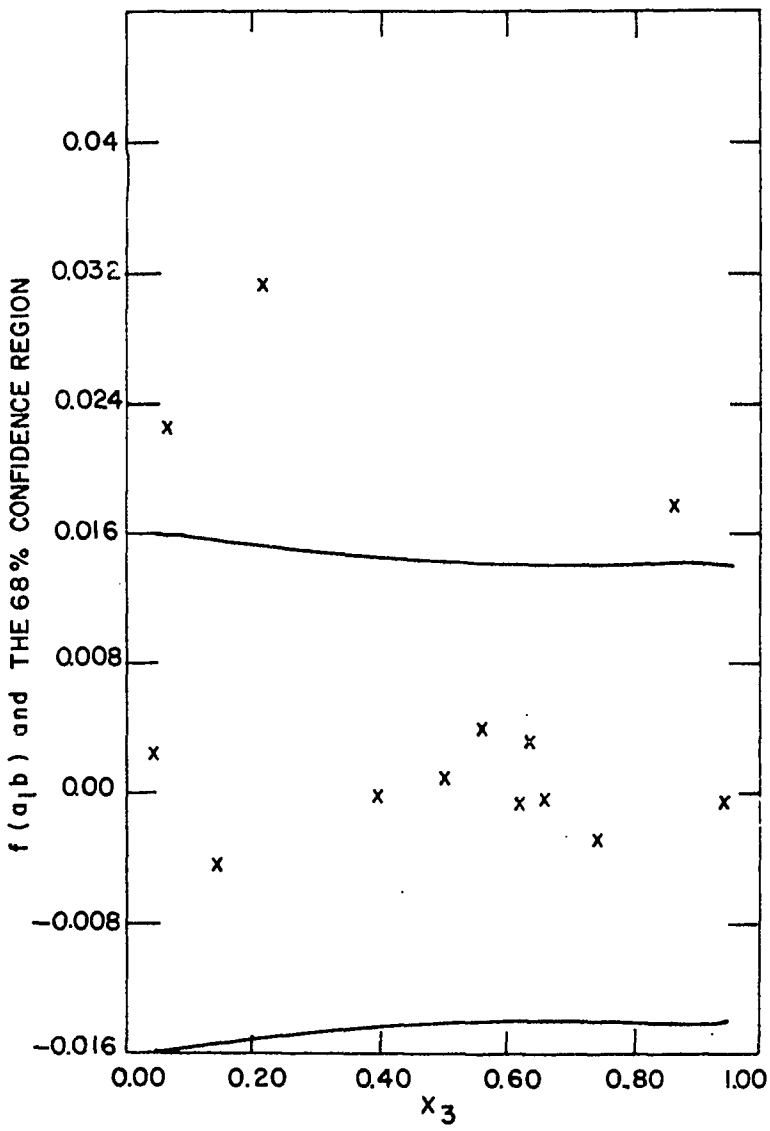

Figure 29. The local area test results for the acetone-methanol system

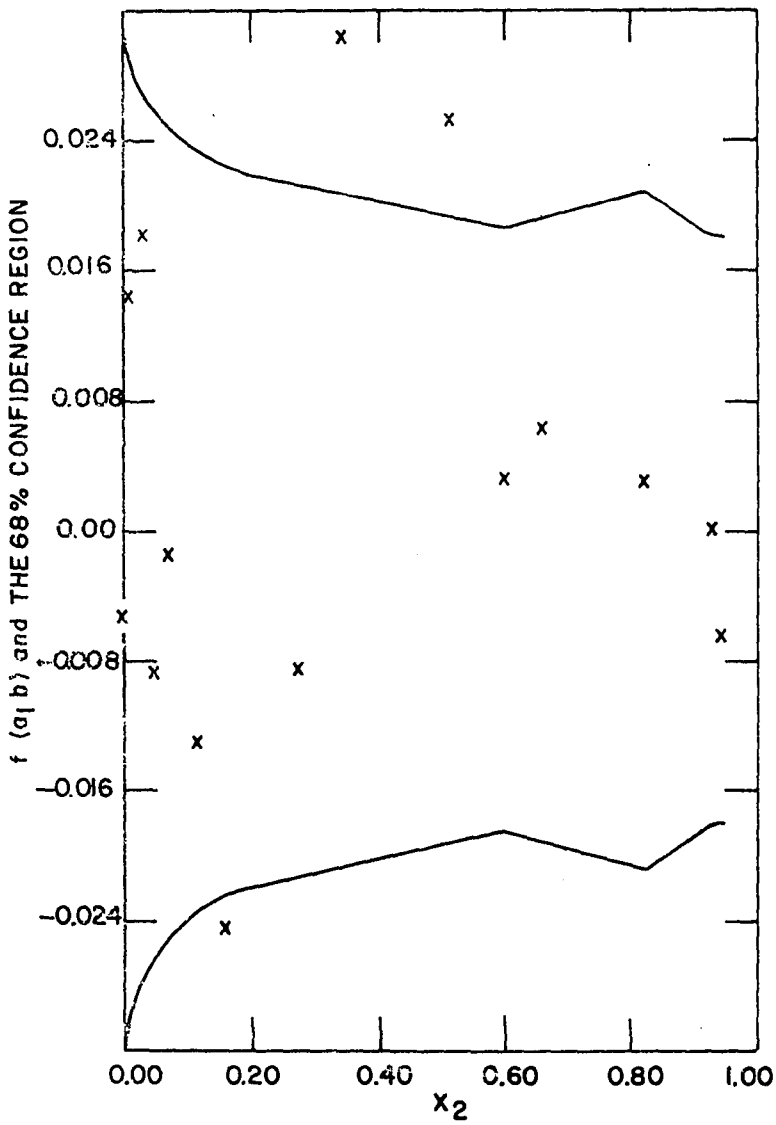

Figure 30. The local area test results for the methanol-water system

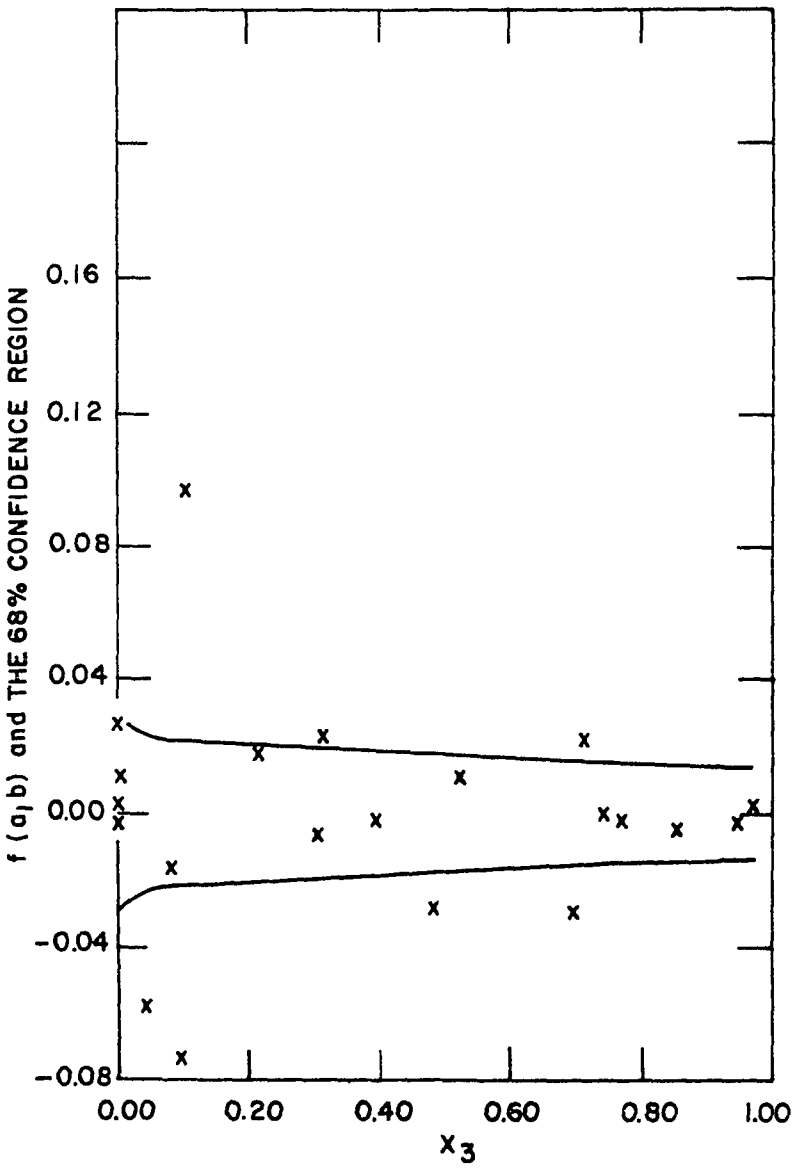

Figure 31. The local area test results for the acetone-water system

The confidence region for the latter system is quite well fixed by the data as shown in Figure 31.

The values of f(a,b) are very well distributed about zero and show no positive or negative trends with composition. The conclusion that the data are free from systematic error and accurate within the estimated uncertainties of the measurement seems very well justified.

The acetone-methanol data are also used to illustrate one of the problems which could arise when the local area test is summed over a concentration interval as suggested in the procedure presented by Tao (49). Equation 20,

$$Q(b) - Q(a) - \int_a^b \ln\gamma_3/\gamma_2 dx_3 = 0, \qquad (20)$$

can be used to illustrate the problem if a and b are taken as the first and last data points, respectively. The first and last data points for the acetone-methanol system are at $x_3 = 0.047$ and 0.977. Letting b and a correspond to these points, equation 20 is the sum of f(x) over all of the intervening intervals. The algebraic sum of the values of f(a,b) reported in Figure 29 is 0.0736. This value is quite large and would cause the data to be rejected as "inconsistent" by the suggested standard of Prausnitz (37), for example. The large value is caused by the three large positive values of f(a,b) shown in Figure 29. The satisfaction of the local area test on a point-by-point basis is therefore much more meaningful than a cumulative integration over several data points. This system illustrates a situation where the cumulative integration would indicate very inaccurate data when, in fact, it appears to be free from systematic error and contains only three intervals where the data are quite inaccurate. This lends

further support to the contention that the words consistent and inconsistent should not be applied to experimental data.

The Margules equations fit these data rather poorly and hardly need be compared to the confidence region for the activity coefficients. The comparison is shown, however, for the acetone-methanol system. The values of $E(\gamma_i^r)$ obtained from the constants reported by Griswold and Wong from the constants obtained by a least-squares fit of $\ln\gamma_3/\gamma_2$, and from the smooth curves shown in Figure 20 are all shown in Figures 32 and 33 for acetone and methanol, respectively. The errors from the Margules equations are large as expected. The errors from the graphically smoothed curve are smaller and appear to be reasonably distributed about the 68 per cent confidence region. The confidence regions may be slightly large but the graphical curve should not be relied upon too heavily to fix the confidence region.

Griswold and Wong presented 50 data points obtained for the ternary system water-methanol-acetone at 100^0C. The data are shown in Table 10. They estimated the ternary Margules constant, C in equation 91, as 1.1605 but apparently did not apply a thermodynamic consistency test.

The constants that Griswold and Wong obtained for the ternary Margules equations are used to calculate values of γ_i^r at every data point. The local area consistency test is then applied to the data by using equation 50, viz.,

$$f(x) \simeq f(a,b) - f_M(a,b).$$

The values for $f_M(a,b)$ are obtained from equation 48 in the form

$$f_M(a,b) = \sum_{i=1}^{3} \tfrac{1}{2}(x_{ia} + x_{ib})\ln(\gamma_{ib}^r/\gamma_{ia}^r) \qquad (92)$$

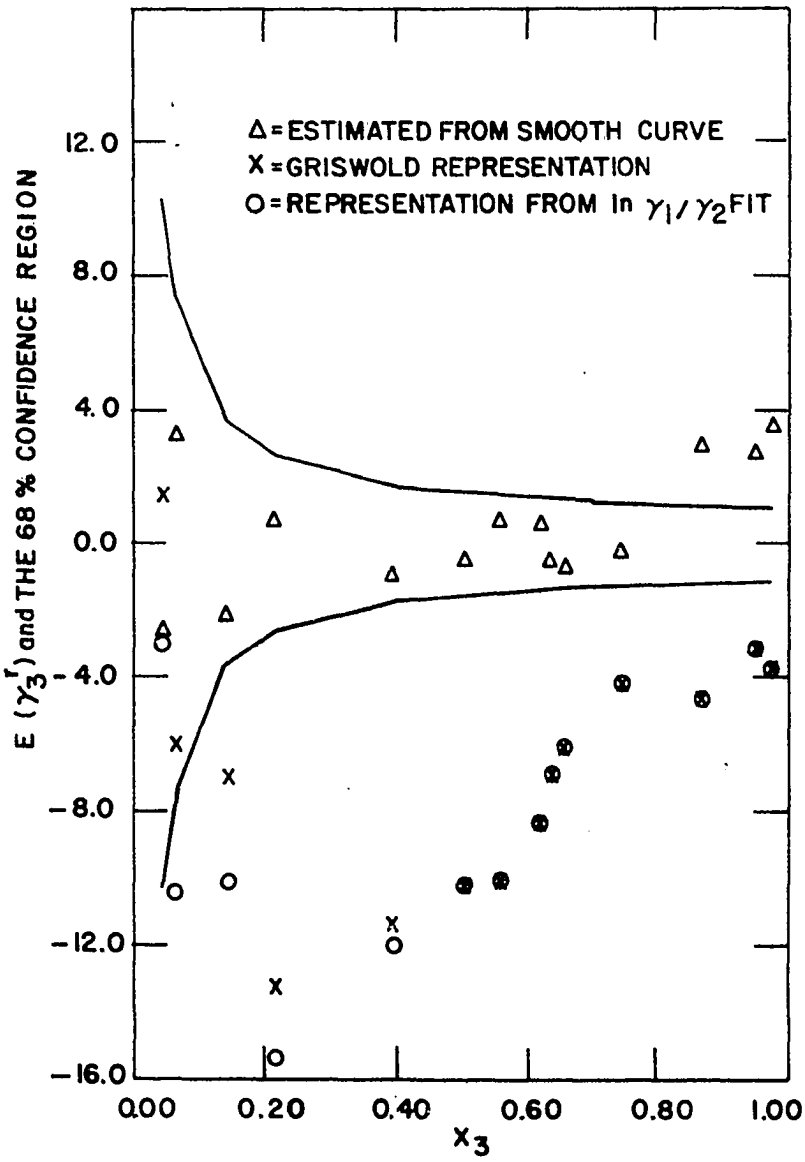

Figure 32. The representation error for the activity coefficient of acetone

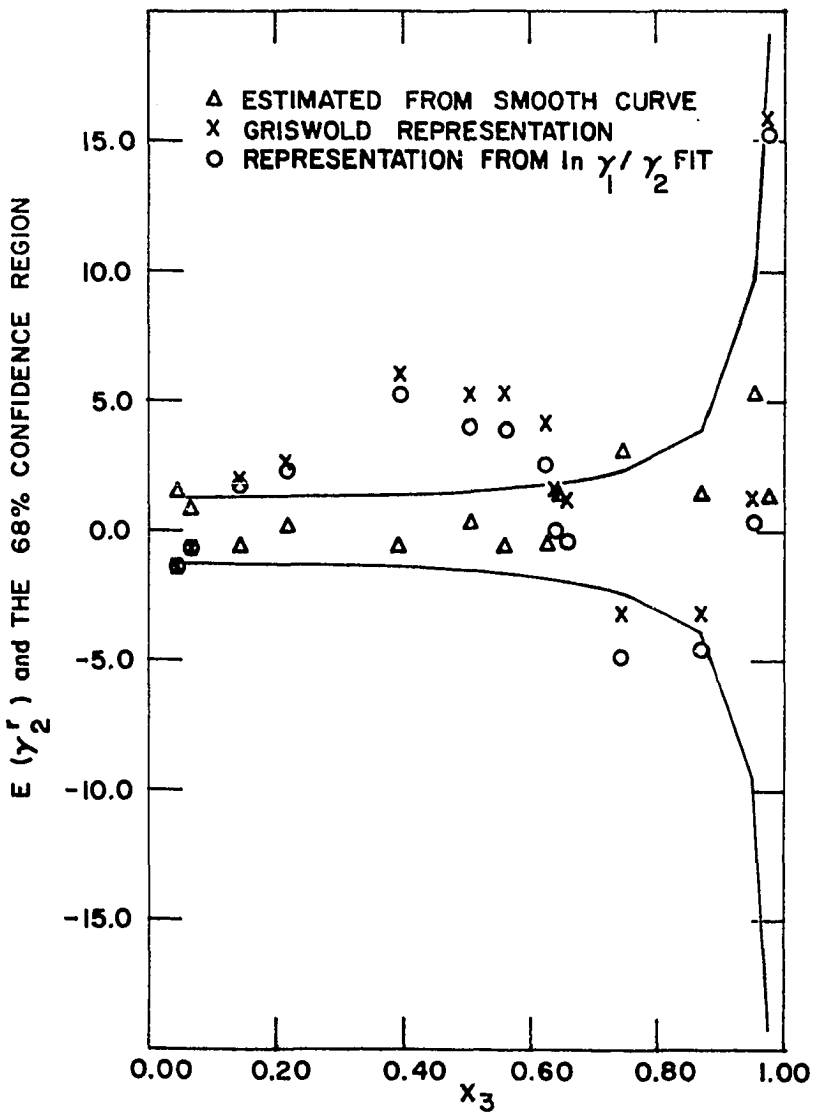

Figure 33. The representation error for the activity coefficient of methanol

Table 10. Vapor-liquid equilibrium data for the water-methanol-acetone system (14) at 100°C[a]

Run No.	x_1	x_2	y_1	y_2	γ_1	γ_2	γ_3	P
1	.051	.679	.029	.648	2.104	1.044	1.228	55.4
2	.034	.050	.031	.069	3.328	1.489	0.996	54.6
3	.063	.330	.040	.339	2.411	1.154	1.076	56.9
4	.070	.549	.045	.525	2.422	1.065	1.176	56.5
5	.103	.640	.055	.621	1.942	1.043	1.272	54.4
6	.130	.639	.065	.624	1.768	1.021	1.324	52.9
7	.136	.479	.071	.474	1.906	1.068	1.197	54.6
8	.169	.654	.084	.657	1.696	1.013	1.391	51.0
9	.202	.616	.099	.629	1.580	1.020	1.407	50.5
10	.212	.294	.112	.295	1.886	1.059	1.190	53.4
11	.250	.549	.122	.558	1.631	1.003	1.483	49.9
12	.264	.546	.126	.535	1.606	0.971	1.672	50.2
13	.171	.059	.128	.064	2.683	1.149	1.045	53.6
14	.302	.478	.143	.489	1.567	1.000	1.544	49.4
15	.246	.082	.160	.077	2.300	0.982	1.116	52.9
16	.322	.299	.163	.296	1.738	1.005	1.363	51.3
17	.381	.344	.190	.339	1.644	0.959	1.574	49.2
18	.369	.171	.194	.145	1.802	0.858	1.370	51.2
19	.416	.440	.197	.495	1.443	1.012	1.820	45.4
20	.455	.470	.220	.593	1.368	1.055	1.978	42.1
21	.426	.130	.222	.084	1.738	0.637	1.453	49.8
22	.505	.243	.242	.229	1.495	0.868	1.829	46.5
23	.512	.127	.247	.087	1.564	0.656	1.670	48.4
24	.552	.345	.263	.443	1.348	1.073	2.265	42.1
25	.546	.287	.263	.301	1.407	0.908	2.147	43.7
26	.608	.160	.279	.135	1.393	0.756	2.143	45.2
27	.571	.338	.283	.442	1.354	1.055	2.317	40.6
28	.609	.213	.285	.199	1.364	0.805	2.363	43.4
29	.752	.033	.302	.041	1.200	1.096	2.560	44.5
30	.639	.233	.306	.267	1.333	0.942	2.606	41.4
31	.754	.088	.319	.092	1.205	0.879	2.976	42.4
32	.707	.193	.328	.261	1.248	1.076	3.108	40.0
33	.762	.119	.341	.156	1.198	1.040	3.188	39.9
34	.733	.198	.358	.317	1.201	1.164	3.263	36.5
35	.780	.131	.365	.167	1.210	0.975	3.825	38.4
36	.732	.225	.369	.419	1.169	1.279	3.243	34.4
37	.763	.204	.381	.411	1.095	1.308	3.906	32.5
38	.815	.119	.391	.197	1.145	1.169	4.198	35.4

[a]Component 1 = water; component 2 = methanol; component 3 = acetone.

Table 10 (Continued)

Run No.	x_1	x_2	y_1	y_2	γ_1	γ_2	γ_3	P
39	.790	.182	.407	.396	1.092	1.367	4.222	31.4
40	.869	.107	.481	.271	1.081	1.467	5.723	28.9
41	.921	.048	.513	.132	1.099	1.609	6.409	29.2
42	.931	.038	.514	.091	1.076	1.382	7.033	28.8
43	.937	.044	.565	.148	1.045	1.730	7.447	25.6
44	.952	.029	.589	.098	1.031	1.673	7.837	24.6
45	.951	.034	.618	.126	1.044	1.768	7.766	23.7
46	.952	.040	.672	.146	1.082	1.663	9.930	22.6
47	.972	.018	.692	.027	1.053	0.660	11.842	21.8
48	.962	.031	.724	.112	1.067	1.525	9.475	20.9
49	.968	.023	.746	.054	1.115	1.009	9.618	21.3
50	.960	.039	.756	.175	1.037	1.716	2.595	19.4

The values for $f_M(a,b)$ would be identically zero if the Margules equations were integrated analytically. However, the error in the trapezoid rule formula causes nonzero values to be obtained from equation 92. If the representation equations accurately describe the system, equation 92 will accurately correct for the error in the trapezoid rule. The effect of using the Margules equations to correct for the length of the integration interval in this manner is investigated while testing the accuracy of the data.

The values of x_1, x_2, y_1, y_2, γ_1, γ_2, γ_3 and P are punched on one computer data card for each data point. The cards are then read in different orders and values of $f(a,b)$, $f_M(a,b)$ and the approximation to $f(x)$ are calculated in the order read. The test is first applied to the data by evaluating $f(x)$ between each pair of points as they appear in Table 10. This is in the general order of increasing

x_1. The results are shown in Figure 34. The 68 per cent confidence region shown is based on the measurement uncertainties estimated for the binaries, viz., $s_{x_1} = 0.004$, $s_{x_2} = 0.004$, $s_{y_1} = 0.004$, $s_{y_2} = 0.004$, $s_p = 0.3$ and $s_T = 0.3$. Since x_3 and y_3 are taken as dependent variables, the values of s_{x_3} and s_{y_3} are 0.00566. The confidence region is drawn by connecting the calculated values of s_f by straight lines. The irregular shape of the curve is due to the existence of particularly large integration intervals in x_2 and x_3 at some points. Thus, the values of s_f reflect the large integration intervals in the same manner as the values of f(a,b). The value of p_f for the 68 per cent confidence region is 30.6 per cent. The value of p_f for the 50 per cent confidence region (not shown) is 24.5 per cent. The confidence region is therefore small. New measurement uncertainties are estimated, using the same order in the data, as $s_{x_1} = s_{x_2} = 0.004$, $s_{y_1} = s_{y_2} = 0.006$, $s_p = 0.3$ and $s_T = 0.8$. Figure 35 shows the values of f(x) and the confidence region for these estimated measurement uncertainties. The values of p_f obtained are 63.4 and 47.0 for the 68 per cent and 50 per cent confidence regions, respectively. The data cards were then carefully shuffled and the data tested again. The confidence region and the values of f(x) both increased but the values of p_f were again 63.4 per cent and 47.0 per cent. The results from still another shuffling were 59.2 per cent and 45.0 per cent for the 68 per cent and 50 per cent confidence regions, respectively. Every time the cards are shuffled in this manner, different points are paired to give a value of f(a,b). Some pairs of points will have compensating errors

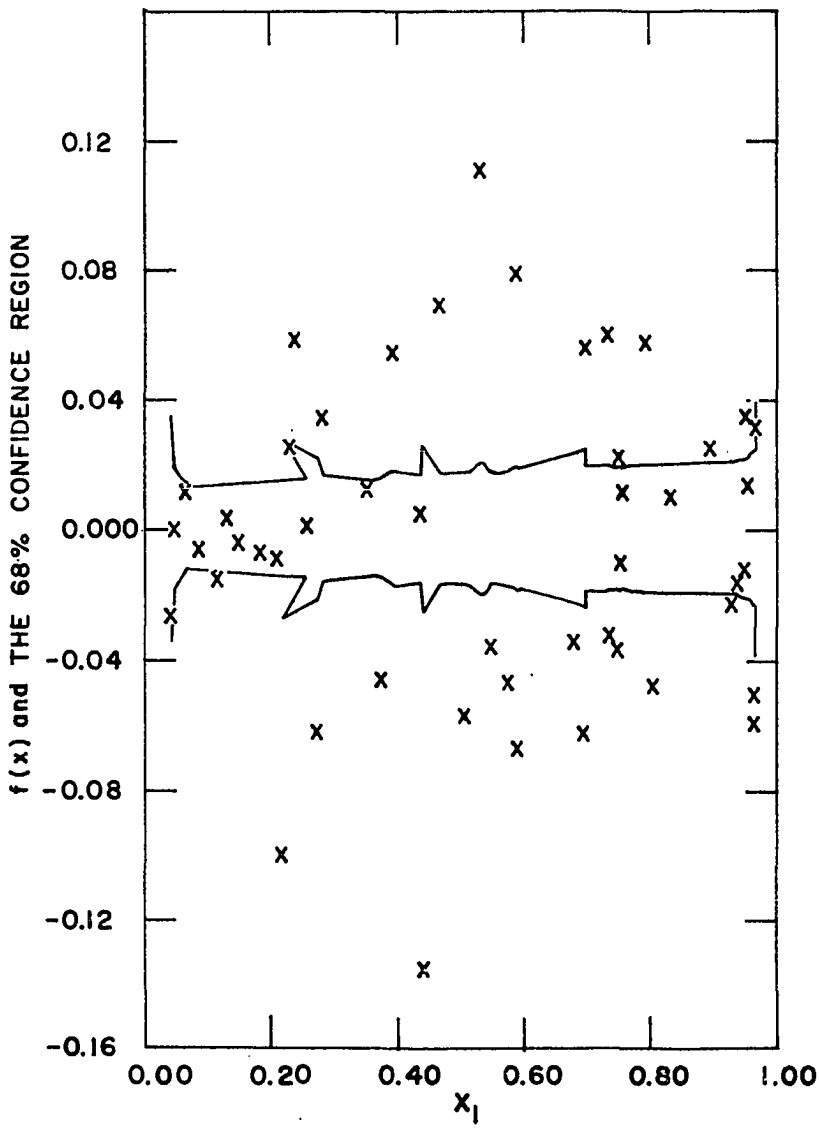

Figure 34. The local area test results for the water-methanol-acetone system with $s_{x_1} = s_{x_2} = 0.004$, $s_{y_1} = s_{y_2} = 0.004$, $s_P = 0.3$ and $s_T = 0.3$

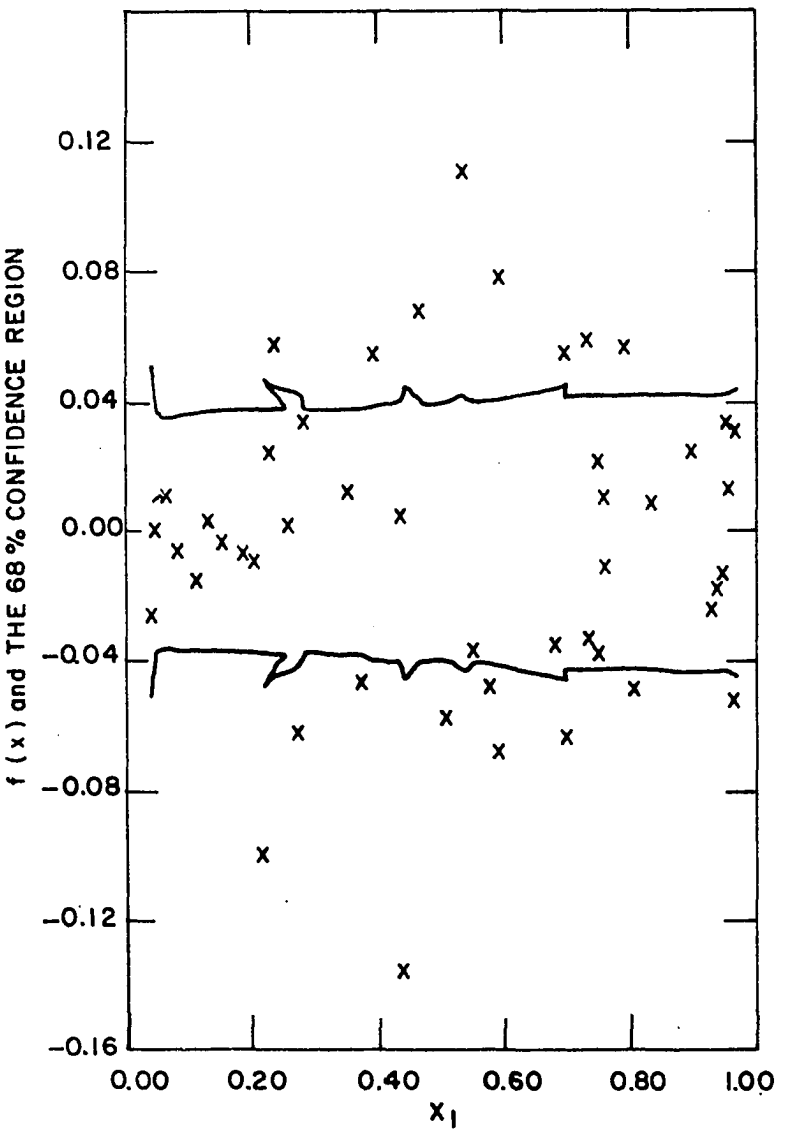

Figure 35. The local area test results for the water-methanol-acetone system with $s_{x_1} = s_{x_2} = 0.004$, $s_{y_1} = s_{y_2} = 0.006$, $s_p = 0.3$ and $s_T = 0.8$

which will give a small value for f(a,b). Other pairs of points will have errors in the opposite directions and will give large values for f(a,b). Therefore, a data point which contains large random error may not appear in every order of integration. In the three different orders of integration reported here, the same eleven data points caused eleven values of f(a,b) to be outside the 68 per cent confidence region in each test. Thus the same eleven data points contributed 22.5 per cent to p_f in each case. There were eighteen additional points which caused values of f(a,b) to exceed s_f in two tests out of three. Ten of these points were common to the first two tests and eight were common to the last two tests. The ten points would contribute 20.4 per cent to p_f for a total of 42.9 per cent out of 63.4 per cent in the first two tests. The integration intervals, in which the eleven points common to all three cases were involved, varied between 0.01 and 0.59 with little apparent affect of the integration interval.

The use of a representation equation to compensate for errors introduced by use of the trapezoid rule appears to be quite effective in permitting a random integration order without concern for the interval of integration. Li and Lu (28) tested portions of the water-methanol-acetone data by arranging the data so that the integration interval was small for each component. They concluded the data for run numbers 29, 31 and 33 in Table 10 were inconsistent. The value of f(a,b) containing point 31, exceeded the value of s_f in all three of the tests made here but points 29 and 33 caused this result only

once each. This procedure, therefore, permits isolating a single point which contains the larger error. There would seem to be little value in isolating such a point when the data contain only random error, however.

The ternary vapor-liquid equilibrium data appear to contain no significant systematic error. The data may be considered accurate within the uncertainties estimated here. The uncertainties are larger than those estimated for the binary data and may indicate both a less careful experimental procedure and more difficult analytical measurements.

The values of $E(\gamma_i^r)$, based on Griswold's representation, and the 68 per cent confidence region are shown in Figure 36 for each of the components. In addition to the lack of fit for the binary system, the ternary Margules constant reported by Griswold and Wong appears to be incorrect.

2. N-octone-ethylbenzene-cellosolve system

Murti and Van Winkle (35) published 71 data points at 760 mmHq for this system. Data for the binary systems n-octane-cellosolve and ethylbenzene-cellosolve at 760 mmHq were reported in the same article. Binary data for the n-octane-ethylbenzene system had been previously reported by Yang and Van Winkle(61). The reported activity coefficients had been corrected for the effect of total pressure by means of virial coefficients. The uncertainties of measurement were not reported.

The activity coefficients for each of the three binary systems

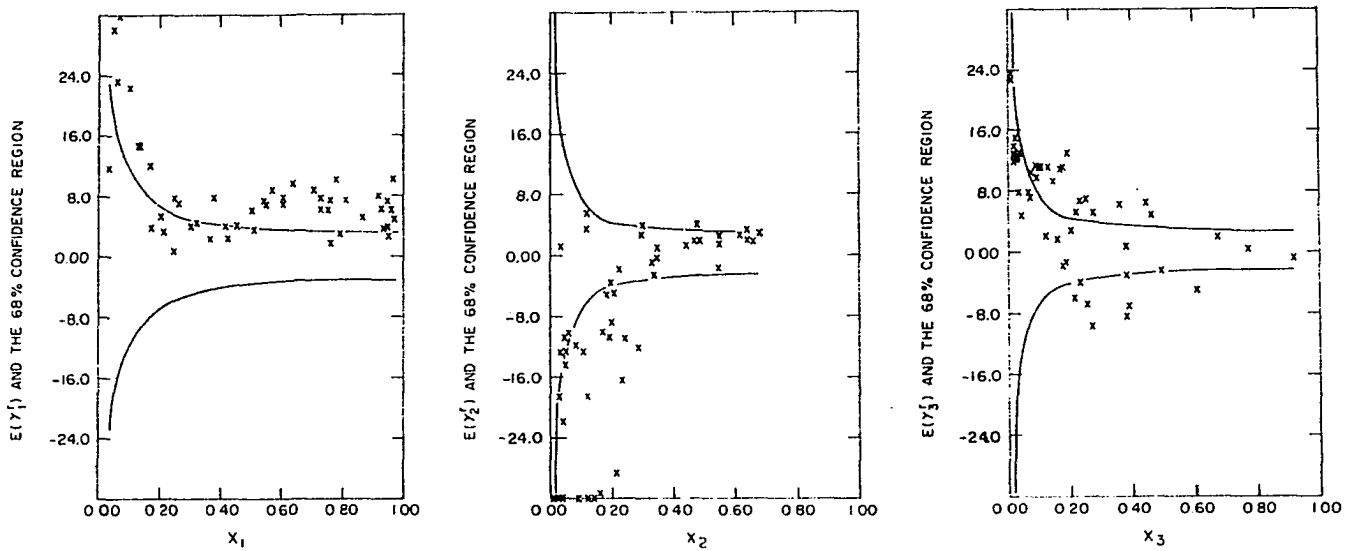

Figure 36. The representation error for the ternary activity coefficients of water, methanol and acetone

had been represented by the Margules equations in the form of equation 90 and the ternary constant of equation 91 had been estimated. The heats of mixing for each of the systems were also measured and reported.

These data are tested in the same manner used for the previous ternary. The heat of mixing is neglected and the binary systems are first tested. Again the same measurement uncertainties are used for each of the three systems. The best estimate of these uncertainties appears to be $s_x = 0.001$, $s_y = 0.001$, $s_p = 0.5$ mmHq and $s_T = 0.3^0$C. The values of f(a,b) and the 68 per cent confidence regions are shown in Figures 37, 38 and 39 for the n-octane-cellosolve, ethylbenzene-cellosolve and n-octane-ethylbenzene systems, respectively. The n-octane-ethylbenzene system appears to be much more accurately measured than the other two systems.

None of the systems show any significant maldistribution about zero or any systematic trends which might indicate the effect of the heat of mixing terms. This may seem somewhat surprising but is explained by the fact that the temperature changes less than 10^0C over the entire composition range for each system. The integral involving ΔH is therefore small over each interval. The maximum value of $(\Delta H/RT^2)(T_b - T_a)$ for the n-octane-cellosolve system occurs at low x_1 and is about 0.0006. The f(a,b) values presented in Figure 37 indicate, therefore, that the heat of mixing integral is negligible.

The ternary system is tested by again using the trapezoid rule integration of the Margules equations to correct for the length of the integration interval. The data were first arranged in order of

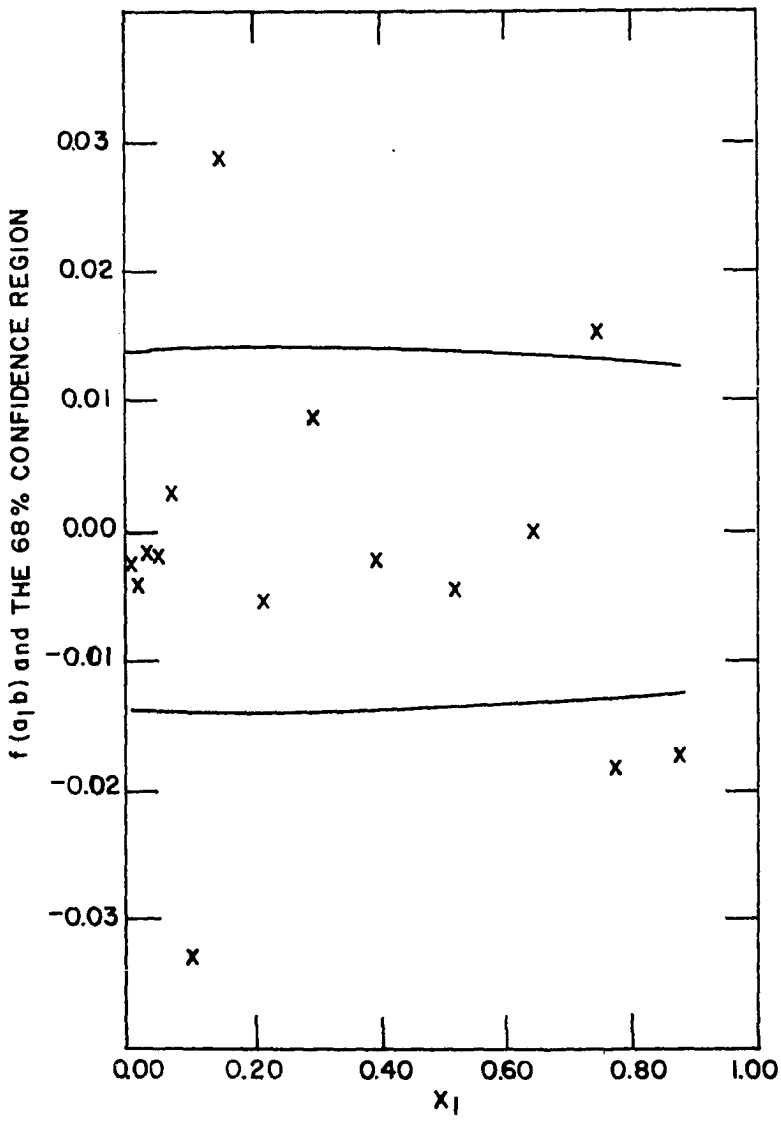

Figure 37. The local area test results for the n-octane-cellosolve system

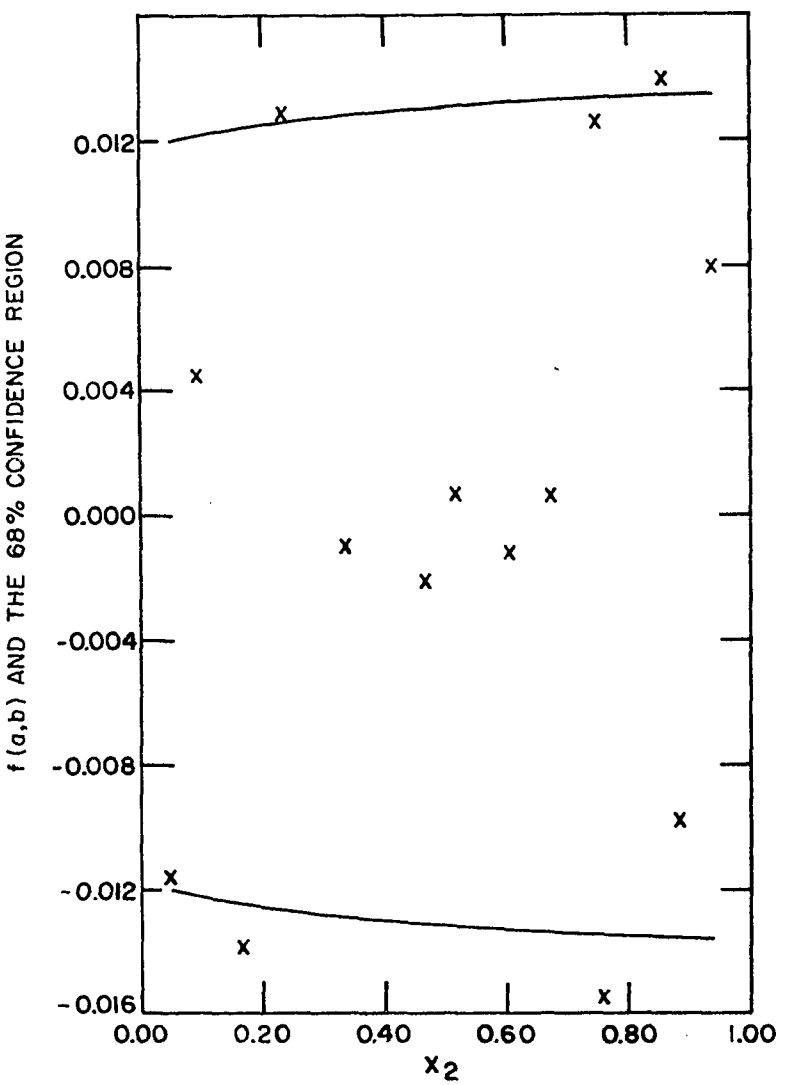

Figure 38. The local area test results for the ethylbenzene-cellosolve system

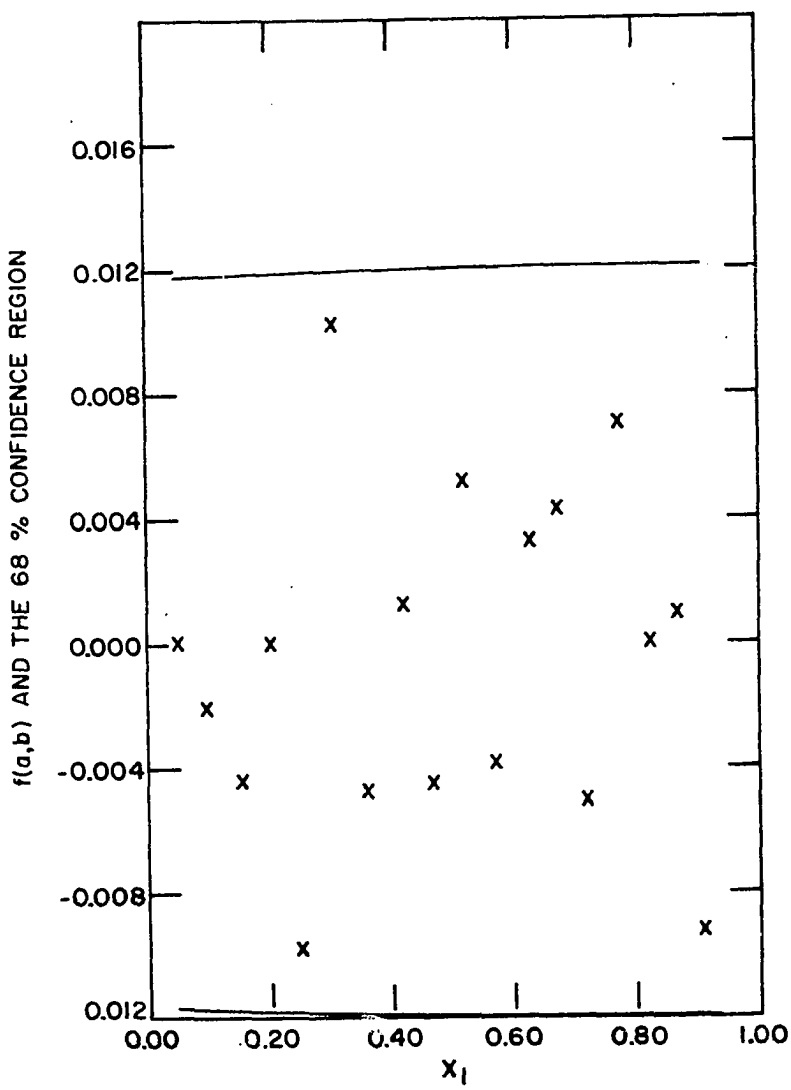

Figure 39. The local area test results for the n-octane-ethylbenzene system

increasing x_1 (n-octane), then in order of increasing x_2 (ethylbenzene) and finally in order of increasing x_3. The results are shown in Figure 40 with f(x) plotted against x_1, x_2 or x_3 depending upon the order of integration. The 68 per cent confidence regions shown are calculated from the uncertainties estimated for the binary systems. The values of p_f for the 68 per cent confidence region are 65.7 per cent, 61.4 per cent and 72.9 per cent for the order in x_1, x_2 and x_3 respectively. The values of p_f for the 50 per cent confidence region are 52.9 per cent, 50.0 per cent and 58.6 per cent in the same order.

These results indicate that the ternary data were obtained with the same accuracy as the binary data and that the heat of mixing for the ternary system is also negligible. The variation among the three values of p_f is apparently caused by a combination of the small confidence region associated with accurately measured data and the lack of fit of the Margules equations. These data indicate that, at least for some systems, ternary data can be obtained as accurately as binary data.

The values of $E(\gamma_i^r)$ are shown in Figure 41 for each component. The error in the activity coefficients is seen to be frequently on the order of 8 per cent.

The procedure previously suggested for estimating the heat of mixing in a binary mixture can not be applied to a ternary system when the order of integration is random as in the examples here. In such a case the heat of mixing must first be estimated and included in the evaluation of f(x). The heat of mixing for this ternary system was

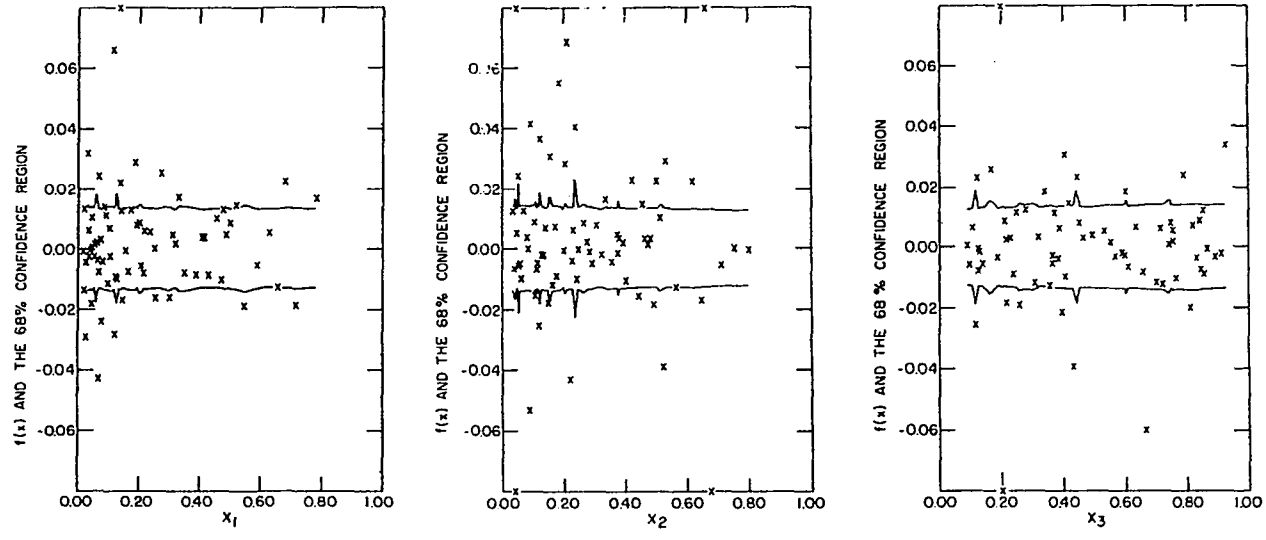

Figure 40. The local area test results for the n-octane-ethylbenzene-cellosolve system

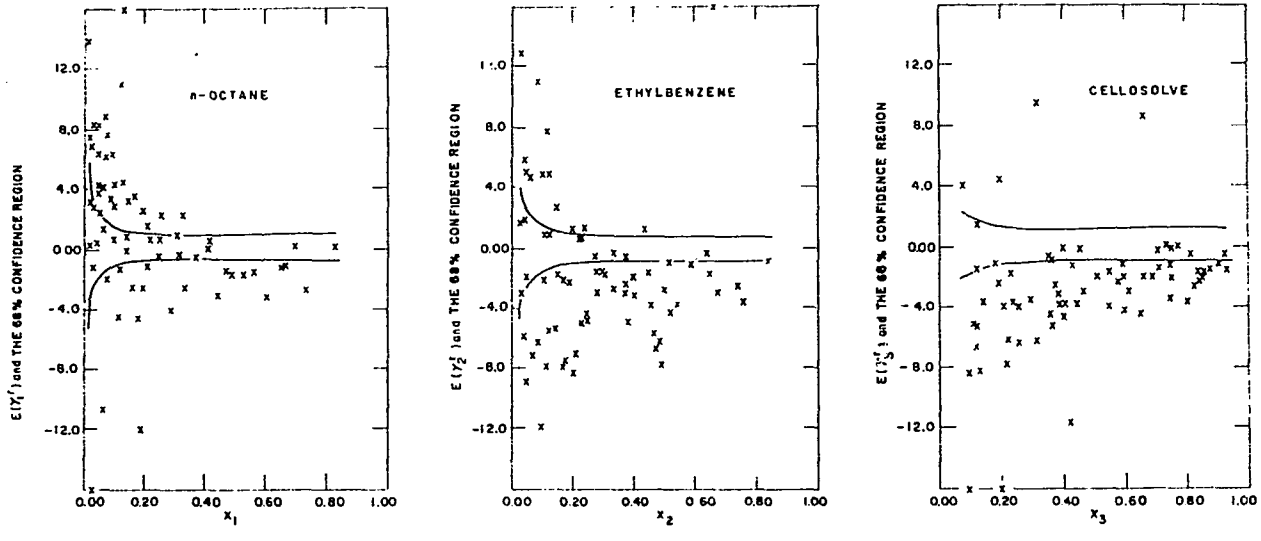

Figure 41. The representation error for the ternary activity coefficients of n-octane, ethylbenzene and cellosolve

reported by Murti and Van Winkle (35). The heat of mixing integral was evaluated for a few intervals and found to be negligible as concluded above.

IV. CONCLUSIONS AND SUMMARY

The thermodynamic consistency tests currently available in the published literature are reviewed. The review indicates that the local area test method as proposed by Stevenson and Sater (45) is the most viable test procedure. This test method indicates local regions of inaccuracy in the data. In addition, the accuracy with which this test can be applied appears to be higher than with any of the other test methods reviewed. The test can be applied directly to the activity coefficient data without the need for smoothing and the relative effects of random and systematic error can therefore be estimated.

Propagation of error equations are used to develop confidence regions for the test results. The confidence regions are calculated from the experimental uncertainties associated with the independently measured variables. If these uncertainties are unknown they can be estimated from the local area test results and a knowledge of the experimental procedure. The maximum uncertainty in any one experimentally measured variable can be estimated or reasonable levels of error in each of the independently measured variables can be obtained if the experimental procedure is not adequately described for a given set of data.

The derived equations for the confidence regions are tested for accuracy by simulating normally distributed random error in the measured variables. The four hypothetical systems described in Tables 2 and 3 are used for this simulation. The confidence regions, based on the assumption of normally distributed random error in the test

results, are shown to be quite accurate by comparing the calculated and theoretical results for these simulations. Tables 4 and 5 show these results. The confidence regions characterize the effect of random error only. Random errors are to be expected in any set of experimental data. The primary purpose of a consistency test is then to detect whether a set of data also contains nonrandom error. Statistical criteria are developed for determining, with some level of confidence, whether nonrandom error also exists.

The words consistent and inconsistent appear to be improperly applied to experimental data because data can never precisely satisfy a mathematical relationship. Thus data should be tested to determine only whether the results are as accurate as can be expected for the given experimental uncertainties. The words consistent and inconsistent should then be reserved to describe representation equations which do or do not, respectively, satisfy the Gibbs-Duhem equation.

The propagation of error equations are also applied to the defining equation for the activity coefficient. The confidence regions for the expected error in the activity coefficient are shown to be accurate by using simulated error in the hypothetical system. The use of these confidence regions in testing the accuracy of representation equations for the ethanol-chloroform system is illustrated. The accuracy of a graphical smoothing of the data is also illustrated.

Phase equilibrium data are frequently tested when the heat of mixing or volume change of mixing is unknown. Methods are proposed and illustrated for estimating the magnitude of these quantities if

the test results indicate the presence of nonrandom error. The accuracy of such a procedure is necessarily low because the test results must first be differentiated.

The application of the test procedure to two ternary systems is illustrated. The experimental measurement uncertainties for the water-methanol-acetone data are shown to be larger than the estimates given by Griswold and Wong (14). The probability is high, however, that the data contain only random error. The ternary data are also shown to contain larger experimental uncertainties than the binary data. The n-octane-ethylbenzene-cellosolve data (35) also appear to contain only random error. The ternary data, in this case, are as accurately measured as the binary data. The Margules equations are shown to represent the data for both ternary systems far less accurately than is justified by the uncertainty of the experimental measurements. However, the representation appears to be sufficiently accurate to correct for the integration error incurred by use of the trapezoid rule. The consistency test is applied to the complete set of data in both cases and little effect of the order of integration is observed.

V. RECOMMENDATIONS

The equations derived here have been shown to be accurate by using simulated systems. Their use has been illustrated by a few applications. However, a detailed comparison of the utility of this test procedure as compared to other test procedures has not been developed. Several sets of published experimental data, containing various types of errors, should be tested by each of the test procedures and the results compared. The publication of such comparisons is needed to improve the general understanding of the power and the limitations of thermodynamic consistency tests.

The effects of constant relative error as opposed to constant absolute error is briefly discussed in this work. The nature of the errors actually encountered in obtaining phase equilibrium data is not considered in depth. The choice among various analytical procedures to be used and the consideration of the optimum experimental design would appear to be logical extensions of this work.

The use of the Gibbs-Duhem equation as an auxiliary relation to reduce the number of experimental measurements required appears to have gained in popularity in recent years. The propagation of error equations should be applied to the equations used in such procedures to indicate the effect of measurement uncertainty on the final result. Such a development could provide a means for determining whether all the variables should be measured or whether one measurement should be deleted and that variable calculated by means of the Gibbs-Duhem equation.

Finally, the effect of measuring N or N - 1 compositions in an N component system should be investigated. This subject is briefly considered in this work but the effect of the alternatives on the consistency test results is not considered. A part of such an investigation would be to consider the various methods of closing a material balance when N compositions, each of which contains random error, are measured.

VI. LITERATURE CITED

1. Barker, J. A. Determination of activity coefficients from total pressure measurements. Australian Journal of Chemistry 6: 207-210. 1953.

2. Beatty, Harold A. and Calingaert, George. Test for the accuracy of vapor-liquid equilibrium data. Industrial and Engineering Chemistry 26: 904-909. 1934.

3. Black, Cline. Phase equilibria in binary and multicomponent systems. Industrial and Engineering Chemistry 50: 403-412. 1958.

4. Black, Cline, Derr, E. L. and Papadopoulos, M. N. Systematic prediction of separation factors. II. Phase equilibria for process design. Industrial and Engineering Chemistry 55: 38-47. 1963.

5. Bourne, J. A. Possible errors in the application of the area test to vapour-liquid equilibria. Chemical Engineering Science 19: 508-509. 1964.

6. Broughton, D. B. and Brearley, C. S. Adjustment of vapor-liquid equilibrium data. Industrial and Engineering Chemistry 47: 838-843. 1955.

7. Carlson, Harrison C. and Colburn, Allan P. Vapor-liquid equilibria of nonideal solutions. Industrial and Engineering Chemistry 34: 581-589. 1942.

8. Chang, Shinn-Der and Lu, Benjamin C.-Y. A general consistency test. International Symposium on Distillation. Part 3. Vapor-liquid equilibrium 22-31. 1969. Brighton, England.

9. Deshpande, A. K. and Lu, Benjamin C.-Y. Extension of binary vapor-liquid equilibrium data. The Canadian Journal of Chemical Engineering 41: 84-85. 1963.

10. Forsythe, George E. Generation and use of orthogonal polynomials for data-fitting with a digital computer. Journal of the Society for Industrial and Applied Mathematics 5, No. 2: 74-87. June 1957.

11. Friend, John, Scheller, William A., and Weber, James H. Calculation of net deviations from consistency in low pressure vapor-liquid equilibrium data. Industrial Engineering Chemistry Process Design and Development 9: 144-148. 1970.

12. Gautreaux, M. F. Jr. and Coates, Jesse. Activity coefficients at infinite dilution. American Institute of Chemical Engineers Journal 1: 496-500. 1955.

13. Gilmont, Roger, Weinmar, Eugene A., Kramer, Franklin, Miller, Eugene, Hashmall, Frank and Othmer, Donald F. Thermodynamic correlation of vapor-liquid equilibria. Industrial and Engineering Chemistry 42: 120-126. 1950.

14. Griswold, John and Wong, S. Y. Phase equilibria of the acetone-methanol-water system from 100°C into the critical region. Chemical Engineering Progress Symposium Series 48, No. 3: 18-34. 1952.

15. Herington, E. F. G. A new method for the interpolation of ternary vapor-liquid equilibrium data. Journal of Applied Chemistry 2: 11-19. 1952.

16. Herington, E. F. G. Symmetrical-area tests for the consistency of vapour-liquid equilibrium data. I. Isothermal data. Journal of Applied Chemistry 18: 285-291. 1968.

17. Herington, E. F. G. Tests for the consistency of experimental isobaric vapour-liquid equilibrium data. Journal of the Institute of Petroleum 37: 457-470. 1951.

18. Herington, E. F. G. A thermodynamic test for the internal consistency of experimental data on volatility ratios. Nature 160: 610-611. 1947.

19. Ho, James C. K., Boshko, O., and Lu, Benjamin C.-Y. Determination of vapor-liquid equilibrium data from total pressure-liquid composition measurements. The Canadian Journal of Chemical Engineering 39: 205-208. 1961.

20. Hsi, Chu and Lu, Benjamin C.-Y. Determination of limiting slopes of binary vapor-liquid equilibrium composition curve. The Canadian Journal of Chemical Engineering 47: 90-91. 1969.

21. Ibl, Norbert V. and Dodge, Barnett F. Note on the Duhem equation. Chemical Engineering Science 2: 120-126. 1953.

22. Jost, W. and Röck, H. A convenient method for the evaluation of vapor-liquid equilibria of binary mixtures. Chemical Engineering Science 3: 17-25. 1954.

23. Klaus, R. L. and Van Ness, H. C. An extension of the spline fit technique and applications to thermodynamic data. American Institute of Chemical Engineers Journal 13: 1132-1136. 1967.

24. Klaus, R. L. and Van Ness, H. C. The orthogonal polynomial representation of thermodynamic excess functions. Chemical Engineering Progress Symposium Series 66, No. 81: 88-104. 1967.

25. Klaus, Ronald L. Mathematical treatment of thermodynamic data for solutions. Unpublished Ph.D. thesis. Troy, New York, Library, Rensselaer Polytechnic Institute. 1967. (Microfilm order no. 67-12,837, University Microfilms, Ann Arbor, Michigan).

26. Krishnamurty, V. V. G. and Rao, C. Venkata. Thermodynamic consistency of ternary vapour-liquid equilibrium data. Journal of Scientific and Industrial Research (India) 14B: 188-189. 1955.

27. Lee, Byung-Ik and Edmister, Wayne C. Isothermal enthalpy difference for saturated liquid mixture via isobaric Gibbs-Duhem equation. American Institute of Chemical Engineers Journal 15: 615-617. 1969.

28. Li, James C. M. and Lu, Benjamin C.-Y. A note on thermodynamic consistency of ternary vapor-liquid equilibrium data. The Canadian Journal of Chemical Engineering 37: 117-120. 1959.

29. Ljunglin, J. J. and Van Ness, N. C. Calculation of vapour-liquid equilibria from vapour pressure data. Chemical Engineering Science 17: 531-539. 1962.

30. Lu, Benjamin C.-Y. Eleven quick tests for thermodynamic consistency. Chemical Engineering 67, No. 20: 105-106. October 3, 1960.

31. Lu, Benjamin C.-Y. and Chang, Shinn-Der. On the thermodynamic consistency test of ternary vapour-liquid equilibrium data. Chemical Engineering Science 24: 1405-1406. 1969.

32. Lu, Benjamin C.-Y. and Jones, Hoskyns K. DeQ. Extrapolation of ternary excess thermodynamic properties. The Canadian Journal of Chemical Engineering 44: 251-254. 1966.

33. Mc Dermott, C. and Ellis, S. R. M. A multicomponent consistency test. Chemical Engineering Science 20: 293-296. 1965.

34. Mixon, F. O., Gumowski, Bogdan and Carpenter, B. H. Computation of vapor-liquid equilibrium data from solution vapor pressure measurements. Industrial and Engineering Chemistry Fundamentals 4: 455-459. 1965.

35. Murti, P. S. and Van Winkle, Matthew. Vapor-liquid equilibria and heat of mixing: n-octane-ethylbenzene-cellosolve system. American Institute of Chemical Engineers Journal 3: 517-522. 1957.

36. Othmer, Donald F., Ricciardi, Louis G. and Thakar, Mahesh S. Composition of vapors from boiling binary systems. Industrial and Engineering Chemistry 45: 1815-1821. 1953.

37. Prausnitz, J. M. Molecular thermodynamics of fluid-phase equilibria. Englewood Cliffs, New Jersey, Prentice-Hall, Inc. c1969.

38. Prausnitz, J. M. and Snider, G. D. Thermodynamic consistency test for multicomponent solutions. American Institute of Chemical Engineers Journal 5: 7S-8S. 1959.

39. Prengle, William H., Jr. and Palm, Gordon F. Determination of bubble points of various pressures for prediction of vapor-liquid equilibria. Industrial and Engineering Chemistry 49: 1769-1774. 1957.

40. Prengle, William H., Jr. and Pike, Martin A. Numerical method for calculating excess free energy and activity coefficients from total pressure measurements. Journal of Chemical and Engineering Data 6: 400-404. 1961.

41. Redlich, Otto and Kister, A. T. Algebraic representation of thermodynamic properties and the classification of solutions. Industrial and Engineering Chemistry 40: 345-348. 1948.

42. Redlich, Otto and Kister, A. T. Thermodynamics of nonelectrolyte solutions--x-y-t relations in a binary system. Industrial and Engineering Chemistry 40: 341-345. 1948.

43. Redlich, Otto, Kister, A. T. and Turnquist, C. E. Thermodynamics of solutions--analysis of vapor-liquid equilibria. Chemical Engineering Progress Symposium Series 48, No. 2: 49-61. 1952.

44. Scatchard, George and Raymond, C. L. Vapor-liquid equilibrium. II. Chloroform-ethanol mixtures at 35, 45 and 55°C. Journal of the American Chemical Society 60: 1278-1287. 1938.

45. Stevenson, F. D. and Sater, V. E. Local thermodynamic consistency of vapor-liquid equilibrium data for binary and multicomponent solutions. American Institute of Chemical Engineers Journal 12: 586-588. 1966.

46. Tao, Luh C. General equations for testing consistency of multicomponent vapor-liquid equilibrium data. Industrial and Engineering Chemistry Fundamentals 1: 119-123. 1962.

47. Tao, Luh C. How to compute binary vapor-liquid equilibrium composition from experimental P-x or T-x data. Industrial and Engineering Chemistry 53: 307-309. 1961.

48. Tao, Luh C. New methods of testing thermodynamic consistency of isobaric, multicomponent vapor-liquid equilibrium data. Industrial and Engineering Chemistry Fundamentals 8: 133-137. 1969.

49. Tao, Luh C. Testing the consistency of vapor-liquid equilibrium data. Industrial and Engineering Chemistry 56: 36-41. 1964.

50. Tao, Luh C. A thermodynamic equation relating equilibrium vapor-liquid compositions and enthalpy differences in isobaric multicomponent systems. American Institute of Chemical Engineers Journal 15: 362-366. 1969.

51. Tao, Luh C. Thermodynamic relationship between integral heat of vaporization and isobaric equilibrium vapor-liquid compositions. American Institute of Chemical Engineers Journal 14: 988-989. 1968.

52. Techo, R. An accurate method for testing the thermodynamic consistency of isothermal binary vapour-liquid equilibrium data. Chemical Engineering Science 18: 27-33. 1963.

53. Tierney, John W. Correlation of vapor-liquid equilibrium data. Industrial and Engineering Chemistry 50: 707-710. 1958.

54. Van Ness, H. C. Classical thermodynamics of non-electrolyte solutions. New York, New York. The Macmillan Co. 1964.

55. Van Ness, H. C. Exact forms of the unrestricted Gibbs-Duhem equation. Chemical Engineering Science 10: 225-228. 1959.

56. Van Ness, H. C. On integration of the coexistence equation for binary vapor-liquid equilibrium. American Institute of Chemical Engineers Journal 16: 18-22. 1970.

57. Van Ness, H. C. Precise testing of binary vapour-liquid equilibrium data by the Gibbs, Duhem equation. Chemical Engineering Science 11: 118-124. 1959.

58. Van Ness, H. C. and Mrazek, R. V. Treatment of thermodynamic data for homogeneous binary systems. American Institute of Chemical Engineers Journal 5: 209-212. 1959.

59. White, N. and Lawson, F. The integration of the Gibbs-Duhem equation for a binary two phase system at constant temperature. Chemical Engineering Science 25: 225-234. 1970.

60. Wilson, Grant, M. Vapor-liquid equilibrium XI. A new expression for the excess free energy of mixing. Journal of the American Chemical Society 86: 127-130. 1964.

61. Yang, C. P. and Van Winkle, Matthew. Vapor-liquid equilibria of subatmospheric pressures: n-octane-ethylbenzene system. Industrial and Engineering Chemistry 47: 293-296. 1955.

VII. ACKNOWLEDGEMENTS

The suggestions and guidance of Dr. F. D. Stevenson during the development of this work are appreciated and gratefully acknowledged. More importantly, the encouragement, understanding and typing received from my wife permitted the completion of an otherwise impossible task.